ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Αναγνώριση συναισθήματος με χρήση τεχνικών βαθιάς μάθησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Χρήστου Φ. Χριστοδούλου

Επιβλέπων: Δημήτρης Ασκούνης
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2023

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Συγγραφέας:
Χρήστος Φ. Χριστοδούλου

Επιβλέπων:
Δημήτριος Ασκούνης
Καθηγητής

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 19η Οκτωβρίου 2023.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
Δημήτριος Ασκούνης Καθηγητής	Ιωάννης Ψαρράς Καθηγητής	Ευάγγελος Μαρινάκης Επίκουρος Καθηγητής

19 Οκτωβρίου 2023

Υπογραφή:
_____

Ημερομηνία:
_____

Περίληψη

Η αναγνώριση συναισθημάτων αποτελεί αντικείμενο έρευνας σε διάφορα επιστημονικά πεδία όπως η ψυχολογία, η κοινωνιολογία και η ιατρική. Η παρούσα διπλωματική εργασία εξετάζει το πρόβλημα της αυτόματης αναγνώρισης επτά συναισθημάτων από εικόνες προσώπου με τη χρήση νευρωνικών δικτύων. Στόχος της εργασίας είναι η ανάπτυξη ενός ελαφρού και ευέλικτου μοντέλου που θα μπορεί να αναγνωρίζει συναισθήματα με μεγάλη ακρίβεια όχι μόνο από μετωπικές εικόνες αλλά και από εικόνες που έχουν ληφθεί υπό γωνία.

Για τον σκοπό αυτό, χρησιμοποιήθηκαν διάφορες αρχιτεκτονικές συνελικτικών νευρωνικών δικτύων και πραγματοποιήθηκαν εκτενείς πειραματισμοί σε δύο σύνολα δεδομένων που περιλαμβάνουν εικόνες προσώπων με ετικέτες συναισθημάτων. Τα αποτελέσματα της εργασίας δείχνουν ότι τα προτεινόμενα μοντέλα βαθιάς μάθησης επιτυγχάνουν υψηλή ορθότητα στην αναγνώριση συναισθημάτων από εικόνες προσώπου και θα μπορούσαν να εφαρμοστούν σε πεδία όπου η ανίχνευση των συναισθημάτων είναι ζωτικής σημασίας.

Λέξεις - κλειδιά

Αναγνώριση Συναισθήματος, Ανάλυση Δεδομένων, Ανάλυση Εικόνων, Βαθιά Μάθηση, Συνελικτικά Νευρωνικά Δίκτυα, Τεχνητά Νευρωνικά Δίκτυα, Εξαγωγή Χαρακτηριστικών Προσώπου, Ταξινόμηση Συναισθημάτων, Μηχανική Μάθηση, Υπολογιστική Όραση, Ανίχνευση Συναισθήματος, Πρόβλεψη Συναισθήματος, Τεχνητή Νοημοσύνη, Επιστήμη Δεδομένων

Abstract

Facial emotion recognition is a subject of research in various scientific fields such as psychology, sociology and medicine. This thesis examines the problem of automatic recognition of seven emotions from facial images using artificial neural networks. The aim of the work is to develop a lightweight and versatile model that will be able to recognise emotions with high accuracy not only from frontal images but also images taken at an angle.

For this purpose, multiple convolutional neural network architectures were used and extensive experiments were performed on two datasets that include face images with emotion labels. The results of this diploma thesis show that the suggested deep learning models achieve high accuracy in emotion recognition from facial images and could be applied to various fields where emotion detection is crucial.

Keywords Emotion recognition, Data Analysis, Image analysis, Deep learning, Convolutional neural networks (CNN), Artificial neural networks (ANN), Facial feature extraction, Emotion classification, Machine learning, Computer vision, Emotion detection, Emotion prediction, Artificial Intelligence (AI), Data Science

Ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Δημήτριο Ασκούνη για την ευκαιρία που μου έδωσε να εκπονήσω την διπλωματική μου εργασία πάνω σε ένα τόσο ενδιαφέρον θέμα. Έπειτα, θα ήθελα να ευχαριστήσω τον υποψήφιο διδάκτορα Λουκά Ηλία για τις πολύτιμες συμβουλές του και την καθημερινή καθοδήγηση του κατά την διάρκεια της έρευνας του αντικειμένου καθώς και της συγγραφής του κώδικα και του κειμένου. Ακόμα, θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου οι οποίοι με στήριζαν και με εμψύχωναν τόσο κατά την διάρκεια των σπουδών μου, όσο και στην εκπόνηση της διπλωματικής μου εργασίας.

Αναγνώριση συναισθήματος με χρήση τεχνικών βαθιάς μάθησης

Χρήστος Φ. Χριστοδούλου
el16753@mail.ntua.gr

19 Οκτωβρίου, 2023

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

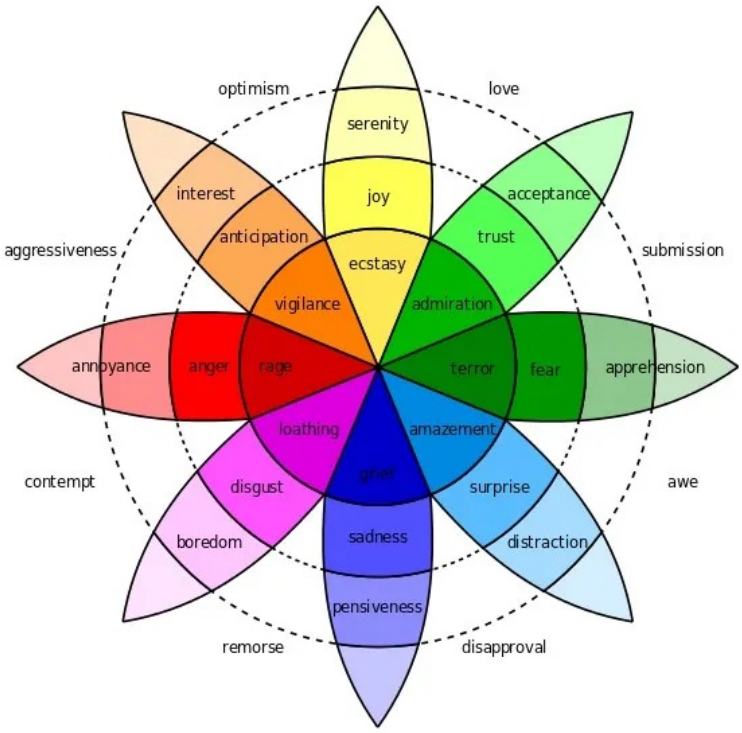
Κεφάλαιο 1

Εισαγωγή

Οι ανθρώπινες εκφράσεις αποτελούν μια μορφή επικοινωνίας δίχως λέξεις και η αποκωδικοποίησή τους αποτελεί αντικείμενο μελέτης εδώ και αρκετά χρόνια [5]. Σύμφωνα με τον Albert Mehrabian [6], κατά την διάρκεια της επικοινωνίας πρόσωπο με πρόσωπο, η έκφραση του προσώπου είναι πιο σημαντική στην μεταβίβαση του μηνύματος σε σχέση με τις λέξεις. Η ανάλυση, κατανόηση και αναγνώριση των ανθρώπινων συναισθημάτων έχει αποτελέσει ένα ζήτημα εξαιρετικού ενδιαφέροντος και έρευνας σε πολλούς επιστημονικούς τομείς, όπως η βιολογία, η ψυχολογία, η κοινωνιολογία και η νευρολογία. Μέσω της έρευνας στους τομείς αυτούς διεξάγονται συμπεράσματα για τις αναπαραστάσεις των συναισθημάτων. Ωστόσο, την λύση στο πρόβλημα της αυτόματης αναγνώρισης μπορεί να δώσει το πεδίο της Υπολογιστικής Όρασης (Computer Vision). Το πεδίο της Υπολογιστικής Όρασης και της Μηχανικής Μάθησης επιδιώκει να αυτοματοποιήσει τη διαδικασία αναγνώρισης με τη χρήση νέων τεχνικών και αλγορίθμων που αναγνωρίζουν και ταξινομούν επιτυχώς τα ανθρώπινα συναισθήματα.

Στην βιβλιογραφία συναντώνται συνήθως έξι συναισθήματα, όπως περιγράφησαν από τους Ekman και Friesen το 1971 [7]. Πιο συγκεκριμένα στις περισσότερες έρευνες συναντώνται ως συναισθήματα η χαρά, η λύπη, ο θυμός, η αηδία, ο φόβος και η έκπληξη. Ο Robert Plutchik το 2001 παρουσίασε την έννοια του τροχού συναισθημάτων, όπου χρωμάτισε τα συναισθήματα και τοποθέτησε τα παρόμοια κοντά και τα αντίθετα αντιδιαμετρικά [1].

Στο πλαίσιο αυτό, η παρούσα διπλωματική εργασία εστιάζει στην πρόβλεψη επτά συναισθημάτων (χαρά, λύπη, θυμός, αηδία, φόβος, έκπληξη και η ουδέτερη έκφραση) από εικόνες προσώπων χρησιμοποιώντας τεχνικές όρασης υπολογιστών. Στόχος της είναι η ανάπτυξη μοντέλων που θα αναγνωρίζουν επιτυχώς με υψηλή ακρίβεια τα συναισθήματα (ή την έλλειψή τους) μέσω των εικόνων και θα δείχνουν τα πιο σημαντικά χαρακτηριστικά που παίζουν ρόλο στην πρόβλεψη. Τέτοια μοντέλα μπορούν αργότερα να εφαρμοστούν σε διάφορους τομείς, όπως η ψυχολογία, η υγεία, τα βιντεοπαιχνίδια, την επαυξημένη πραγματικότητα και το μάρκετινγκ. Για την



Σχήμα 1.1: Χρωματικός τροχός του Plutchik [1]

ακρίβεια, αλγόριθμοι αναγνώρισης συναισθήματος χρησιμοποιούνται για κλινικές διαγνώσεις, για την βελτίωση της επικοινωνίας ανθρώπου και υπολογιστή, ή ακόμα και για αναγνώριση των αντιδράσεων των καταναλωτών ως απόκριση σε μια διαφήμιση.

1.1 Συνεισφορά Διπλωματικής Εργασίας

Στην παρούσα διπλωματική εργασία προτείνονται διάφορες αρχιτεκτονικές συνελικτικών τεχνητών νευρωνικών δικτύων CNN οι οποίες, με ελάχιστη προεπεξεργασία στις εικόνες, προβλέπουν τα διάφορα συναισθήματα των προσώπων. Πιο συγκεκριμένα, εστιάζουμε στο πλήθος και στο μέγεθος των στρωμάτων σε κάθε δίκτυο, καθώς και στην χρήση διάφορων τεχνικών που χρησιμοποιούνται κατά την εκπαίδευση, όπως το Dropout ή το Early Stopping.

Η εργασία αυτή έχει ως στόχο να προτείνει και να αξιολογήσει μια μεθοδολογία για την αναγνώριση πολλαπλών συναισθημάτων σε εικόνες προσώπων. Κατά την ανάπτυξη των αρχιτεκτονικών CNN, λαμβάνονται υπόψη οι απαιτήσεις υπολογιστικής ισχύος και η απόδοση του δικτύου, με στόχο την επίτευξη υψηλής ακρίβειας και αποδοτικότητας στην αναγνώριση συναισθημάτων. Η μεθοδολογία περιλαμβάνει τα ακόλουθα βήματα: προεπεξεργασία των εικόνων προσώπων, εκπαίδευση του δικτύου χρησιμοποιώντας ένα σύνολο δεδομένων που περιλαμβάνει ετικέτες συναισθημάτων, και την αξιολόγηση της απόδοσης του δικτύου σε ένα ανεξάρτητο σύνολο

ελέγχου.

Τα αποτελέσματα και οι αξιολογήσεις της μεθοδολογίας παρουσιάζονται και συζητούνται λεπτομερώς, επισημαίνοντας τα πλεονεκτήματα, τους περιορισμούς και τις πιθανές κατευθύνσεις για μελλοντική έρευνα στον τομέα της αναγνώρισης συναισθημάτων σε εικόνες προσώπων με τη χρήση τεχνητών νευρωνικών δικτύων.

1.2 Δομή κειμένου

Στο παρόν Κεφάλαιο, ως πρώτο, δόθηκε μια γενική εισαγωγή για το αντικείμενο της διπλωματικής και τον τομέα στον οποίο εστιάζει. Στο δεύτερο Κεφάλαιο θα δοθεί το θεωρητικό υπόβαθρο στην τεχνητή νοημοσύνη και την υπολογιστική όραση, με σκοπό να κατανοηθεί καλύτερα οτιδήποτε παρουσιαστεί στην συνέχεια. Έπειτα, στο Κεφάλαιο 3, θα παρουσιαστεί η βιβλιογραφία που προϋπάρχει στον τομέα της αναγνώρισης συναισθημάτων. Στο Κεφάλαιο 4, θα δοθεί μια επισκόπηση πάνω στις τεχνικές λεπτομέρειες που αφορούν το προγραμματιστικό περιβάλλον που χρησιμοποιήθηκε για την παραγωγή τον αποτελεσμάτων.

Στο Κεφάλαιο 5, θα περιγράψουμε τα διάφορα datasets εικόνων που εισήχθησαν στα μοντέλα μηχανικής μάθησης. Στο έκτο Κεφάλαιο θα παραθέσουμε την μεθοδολογία που ακολουθήθηκε για την υλοποίηση των μοντέλων μηχανικής μάθησης που προβλέπουν συναισθήματα από εικόνες προσώπων. Στο έβδομο κεφάλαιο θα παρουσιαστούν τα αποτελέσματα των προαναφερθέντων μοντέλων. Τέλος, στο Κεφάλαιο 8, θα παρατεθούν τα συμπεράσματα της διπλωματικής εργασίας, αλλά και προτάσεις για θέματα που χρίζουν περεταίρω διερεύνησης στο μέλλον.

Κεφάλαιο 2

Τεχνητή Νοημοσύνη και Υπολογιστική Όραση

2.1 Τεχνητή Νοημοσύνη

Η τεχνητή νοημοσύνη (στα Αγγλικά Artificial Intelligence ή εν συντομία AI) είναι ένα πεδίο της επιστήμης υπολογιστών. Ασχολείται με την δημιουργία λογισμικού, καθώς και υλικού, το οποίο μπορεί να φτάσει ή και να ξεπεράσει, κατά κάποιον τρόπο, τις ανθρώπινες νοητικές ικανότητες, εξού και το όνομα της.

Ο όρος "Τεχνητή Νοημοσύνη" φαίνεται να επινοήθηκε από τον μαθηματικό Τζον Μακάρθι (John McCarthy), γνωστό και ως "Πατέρα της Τεχνητής Νοημοσύνης", ο οποίος το 1955 την όρισε ως "την επιστήμη και μηχανική της κατασκευής έξυπνων μηχανών" [8]. Ωστόσο, μπορούμε να πούμε πως τα πρώτα θεμέλια τέθηκαν από τον πρωτοπόρο μαθηματικό Άλαν Μάθισον Τούρινγκ. Το 1936, ο Τούρινγκ πρώτος περιέγραψε μια υποθετική μηχανή υπολογισμών, γνωστή σήμερα και ως "Μηχανή Τούρινγκ" [9]. Επίσης, πρότεινε την ιδέα μιας μηχανής που μπορεί να προσομοιώσει μια άλλη μηχανή, προτείνοντας ουσιαστικά έναν υπολογιστή που μπορεί να προγραμματιστεί.

Το 1950, πρότεινε τη γνωστή Δοκιμή Τούρινγκ (Turing Test) ή "Παιχνίδι της μίμησης" (Imitation game), με την οποία θα μπορεί κανείς να αναγνωρίσει αν μια μηχανή έχει νοημοσύνη ισοδύναμη με ενός ανθρώπου. Σύμφωνα με τον Τούρινγκ, ο στόχος για να περάσει το τεστ επιτυχώς είναι μια μηχανή (ένας υπολογιστής) να μπορεί να συμμετάσχει σε διαλόγους χρησιμοποιώντας ανθρώπινη γλώσσα και να απαντήσει σε διάφορα ερωτήματα, χωρίς ο συνομιλητής να καταλάβει πως πρόκειται για μηχανή [10].

Τα πρώτα συστήματα τεχνητής νοημοσύνης φαίνεται να αναπτύχθηκαν στα μέσα του 20ου αιώνα. Κάποια από αυτά περιλαμβάνουν το "Logic Theorist" [11], το οποίο αναπτύχθηκε με σκοπό να αποδεικνύει μαθηματικά θεωρήματα και το γνώριμο

"ELIZA", που προσομοίωνε έναν διάλογο με έναν άνθρωπο, χρησιμποιώντας τεχνικές αντιστοίχισης μοτίβων. Το "ELIZA" θεωρείται ένα από τα πρώτα chatbots, το οποίο επιστράτευε τεχνικές επεξεργασίας φυσικής γλώσσας (Natural Language Processing).

Ακολούθησαν πολλά συστήματα που επιδείκνευαν ικανότητες όπως το να παίζουν σκάκι, να συνομιλούν με ανθρώπους, να αναγνωρίζουν πρόσωπα, και να διαβάζουν κείμενα. Αξίζει να σημειωθεί πως το Deep Blue της IBM έγινε το πρώτο σύστημα που κέρδισε αγώνα σκακιού υπό χρονικούς περιορισμούς, εναντίον ενός παγκοσμίου πρωταθλητή. Φυσικά δεν θα μπορούσαν να λείπουν εφαρμογές της ΤΝ με φυσική υπόσταση, όπως ρομπότ που πλοηγούνταν στον χώρο.

Όπως είναι φυσικό, το πεδίο αυτό προχώρησε και εξελίχθηκε, προσφέροντας αμέτρητες δυνατότητες και ευκαιρίες στον κόσμο μέχρι σήμερα. Εφαρμογές της Τεχνητής Νοημοσύνης υπάρχουν στην πλειοψηφία των ψηφιακών συστημάτων που χρησιμοποιούνται τη σήμερον ημέρα. Πολλές φορές η Τεχνητή Νοημοσύνη υπάρχει σιωπηλά σε αντικείμενα ή λογισμικά που χρησιμοποιούμε, χωρίς να το αντιλαμβανόμαστε.

Τρανό παράδειγμα είναι τα κινητά τηλέφωνα και οι αμέτρητες "έξυπνες" λειτουργίες τους, όπως το να αναγνωρίζουν τους χρήστες μέσω των χαρακτηριστικών του προσώπου ή του δακτυλικού τους αποτυπώματος ή το να μετατρέπουν την φωνή σε κείμενο. Άλλο παράδειγμα είναι τα λογισμικά πλοήγησης (Global Positioning Systems - GPS) τα οποία αναλύουν την κίνηση σε πραγματικό χρόνο και προτείνουν την βέλτιστη διαδρομή, ελαχιστοποιώντας τον χρόνο ταξιδίου ή ακόμα και την κατανάλωση καυσίμων.

Τέλος, δεν θα μπορούσε να λείπει από τα παραδείγματα το πλέον πασίγνωστο και ευρέως χρησιμοποιούμενο Chat GPT [12]. Το Chat GPT, που από κάποιους θεωρείται chatbot, είναι ένα μοντέλο γλώσσας Τεχνητής Νοημοσύνης (AI language model) με δυνατότητες όπως η παραγωγή και κατανόηση φυσικής γλώσσας, η μετάφραση κειμένων, η περίληψη κειμένων, καθώς και διάφοροι υπολογισμοί, πολύπλοκοι ή μη.

2.2 Πεδία της Τεχνητής Νοημοσύνης

2.2.1 Μηχανική Μάθηση

Ο τομέας της Τεχνητής Νοημοσύνης περιλαμβάνει διάφορους πιο εξειδικευμένους τομείς. Οι τομείς που θα αναφερθούν περιλαμβάνουν ο καθένας ένα διαφορετικό κομμάτι της ΤΝ, χωρίς να σημαίνει όμως αυτό ότι δεν επικαλύπτονται και δεν μοιράζονται διάφορα στοιχεία μεταξύ τους.

Ο πιο γνωστός τομέας της Τεχνητής Νοημοσύνης λέγεται Μηχανική Μάθηση (Machine Learning) και περιλαμβάνει την δημιουργία αλγόριθμων και μοντέλων με σκοπό την πρόβλεψη και λήψη αποφάσεων, βασιζόμενοι σε δεδομένα. Τεχνικές μηχανικής

μάθησης χρησιμοποιούνται όχι μόνο για να προβλέψουν διάφορα γεγονότα (όπως πχ τα spam emails, τις απάτες σε τραπεζικές συναλλαγές, τον καιρό, σεισμούς κ.α.) αλλά και για να βρουν υπο-ομάδες ή μοτίβα μέσα σε τεράστιους όγκους δεδομένων, όπου το ανθρώπινο μυαλό θα χρειαζόταν μέρες ή και μήνες για να κατανοήσει και να αναλύσει.

Τα δεδομένα (data) που εισάγονται σε ένα μοντέλο μηχανικής μάθησης με σκοπό μια πρόβλεψη ή τον διαχωρισμό των δεδομένων σε ομάδες είναι από τους πιο σημαντικούς παράγοντες για την καλή του απόδοση. Μάλιστα στην επιστήμη των υπολογιστών υπάρχει μια φράση η οποία δηλώνει "Garbage in, garbage out", η οποία θα μπορούσε να μεταφραστεί στα Ελληνικά ως "ο,τι δίνεις παίρνεις". Η φράση αυτή τονίζει την σημασία της σωστής εισόδου σε ένα υπολογιστικό σύστημα. Υπό την έννοια της μηχανικής μάθησης, τονίζει την σημασία των σωστών δεδομένων που θα εισαχθούν σε έναν αλγόριθμο. Αν εισάγουμε, για παράδειγμα, δεδομένα με πολλαπλές πανομοιότυπες παρατηρήσεις, δεδομένα ελλειπή ή μη κανονικοποιημένα (στα αγγλικά συναντάμε συχνά τον όρο "dirty" data), είναι πολύ πιθανό να λάβουμε έξοδο που θα είναι εξίσου προβληματική. Είναι, επομένως, σημαντικό για την ποιότητα των δεδομένων εξόδου να εισάγουμε δεδομένα "καθαρά" ("clean" data) και ποιοτικά.

Τα δεδομένα εισόδου (input data) θα πρέπει συνήθως να είναι σε μορφή πίνακα, όπου κάθε γραμμή εκφράζει μια παρατήρηση ή αλλιώς δείγμα (sample) και κάθε στήλη εκφράζει ένα χαρακτηριστικό (feature ή variable). Επιπροσθέτως, θα πρέπει τα δεδομένα να είναι καταλλήλως προ-επεξεργασμένα (pre-processed) σε μια μορφή αποδεκτή από τους αλγόριθμους μηχανικής μάθησης.

Πολύ συχνά, ιδιαίτερα όταν σκοπός είναι ένα μοντέλο πρόβλεψης, τα δεδομένα χωρίζονται σε αυτά που θα αξιοποιηθούν για την εκπαίδευση του αλγόριθμου (training data) και σε αυτά που θα χρησιμοποιηθούν για να αξιολογηθεί το μοντέλο (test data). Τα πρώτα χρησιμοποιούνται για να "εκπαιδευτεί" και να μάθει το μοντέλο τα δεδομένα εισόδου όσο καλύτερα γίνεται, ενώ τα δεύτερα θα εισαχθούν στο μοντέλο για να μας δώσουν μια εικόνα της απόδοσης του. Είναι σημαντικό το μοντέλο να μην έχει δει τα test data, έτσι ώστε να είναι όσο πιο αμερόληπτο γίνεται κατά τη διάρκεια της αξιολόγησης και να μην υπάρξει το λεγόμενο data leakage.

Το "data leakage" (διαρροή δεδομένων) είναι μια κατάσταση που μπορεί να προκύψει όταν δεδομένα που έχουν χρησιμοποιηθεί για την εκπαίδευση ενός μοντέλου εμφανίζονται και πάλι στο σύνολο ελέγχου. Αυτό μπορεί να οδηγήσει σε ανεπιθύμητα αποτελέσματα κατά την αξιολόγηση της απόδοσης του μοντέλου, όπως σε μια υπερεκτίμηση της απόδοσης των μοντέλων, διότι τα μοντέλα θα είναι ήδη εξοικειωμένα με τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευσή τους.

Οι μέθοδοι αξιολόγησης ενός μοντέλου ποικίλουν και πρόκειται να αναλυθούν στην συνέχεια.

2.2.2 Κατηγορίες μηχανικής μάθησης

Η μηχανική μάθηση χωρίζεται σε τρεις βασικές υπο-κατηγορίες:

- **Επιβλεπόμενη μάθηση (Supervised Learning):** Η επιβλεπόμενη μάθηση ονομάζεται έτσι καθώς ο αλγόριθμος κατά τη διάρκεια της εκπαίδευσής του δέχεται ως είσοδο και την ετικέτα ή στόχο (label ή target) του κάθε δείγματος. Για παράδειγμα, για να χτίσει κανείς ένα μοντέλο που θα διακρίνει spam e-mails, θα πρέπει να δώσει και την πραγματική ετικέτα του κάθε email, δηλαδή αν είναι spam ή όχι. Όσο το μοντέλο εκπαιδεύεται, μαθαίνει να ξεχωρίζει τα δείγματα βάσει των χαρακτηριστικών τους, βρίσκοντας όσο τον δυνατόν καλύτερα τους γενικούς κανόνες που διακρίνουν τις δύο ετικέτες.

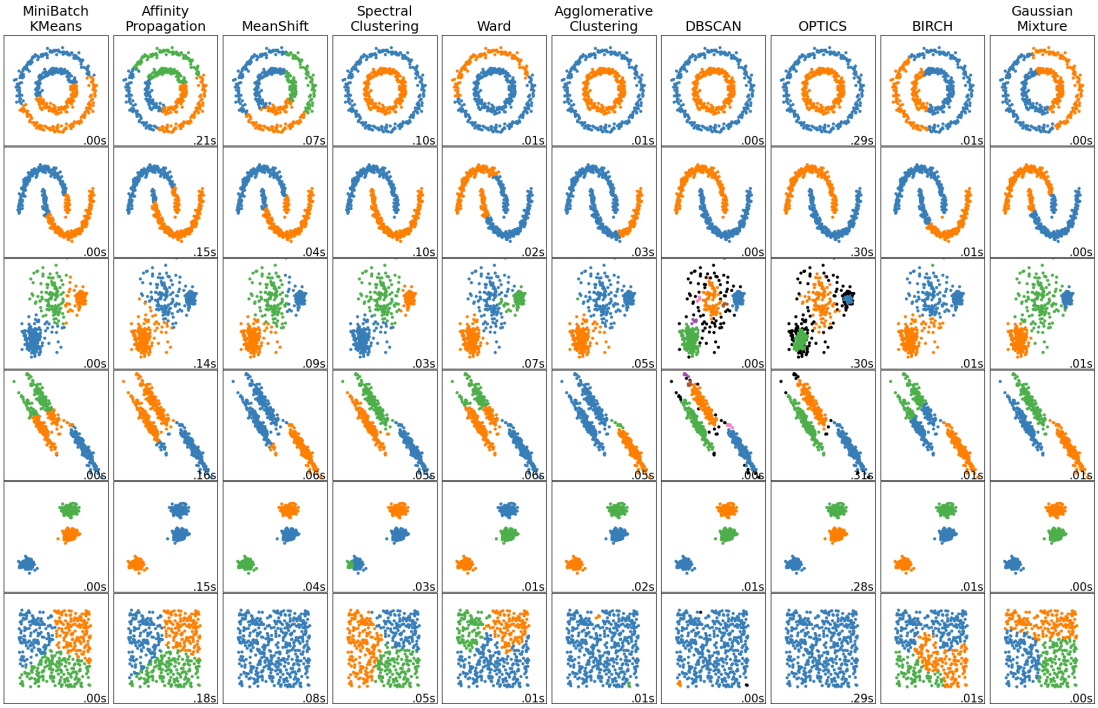
  Οι αλγόριθμοι που ανήκουν στην κατηγορία της επιβλεπόμενης μάθησης διακρίνονται επίσης σε δύο υποκατηγορίες, ανάλογα με τον τύπο της ετικέτας.

  - **Αλγόριθμοι Ταξινόμησης (Classification Algorithms):** Σε αυτή την υποκατηγορία, ο στόχος της πρόβλεψης είναι μια ετικέτα διακριτή, συνήθως με την μορφή λέξης. Αν ο στόχος έχει δύο πιθανές ετικέτες, τότε έχουμε δυαδική ταξινόμηση (binary classification), ενώ αν έχει παραπάνω από δύο τότε έχουμε ταξινόμηση πολλαπλών κλάσεων (multi-class classification). Αν, δε, μια παρατήρηση μπορεί να έχει παραπάνω από μία ετικέτες, τότε έχουμε το λεγόμενο multi-label classification.

    Κλασικά παραδείγματα ταξινόμησης είναι η πρόβλεψη της ενοχλητικής αλληλογραφίας, η πρόβλεψη μιας ασθένειας όπως ο καρκίνος, ο εντοπισμός ενός αντικειμένου σε μια εικόνα (object detection), η πρόβλεψη μιας κακόβουλης ή μη τραπεζικής συναλλαγής, ή ακόμα και η πρόβλεψη του συναισθήματος ενός προσώπου. Το τελευταίο, που είναι και το θέμα της παρούσας διπλωματικής εργασίας, είναι μια περίπτωση multi-class classification, καθώς για κάθε εικόνα μπορεί να προβλεφθεί ένα και μόνο συναίσθημα.

  - **Αλγόριθμοι Παλινδρόμισης (Regression Algorithms):** Στην υποκατηγορία της παλινδρόμησης ανήκουν οι αλγόριθμοι που έχουν ως στόχο πρόβλεψης μία συνεχή μεταβλητή. Εδώ η πρόβλεψη, δηλαδή, δεν κλείνεται σε ένα προκαθορισμένο σύνολο αλλά μπορεί να πάρει οποιαδήποτε συνεχή τιμή. Παραδείγματα παλινδρόμισης περιλαμβάνουν την πρόβλεψη βάρους, ύψους, θερμοκρασίας, τιμής κ.α. Πολλές φορές μοντέλα παλινδρόμισης χρησιμοποιούνται στο πεδίο των οικονομικών επιστημών, με απώτερο σκοπό να δώσουν φως στις τάσεις της αγοράς (market forecasting).

- **Μη επιβλεπόμενη μάθηση (Unsupervised Learning):** Στην περίπτωση της μη επιβλεπόμενης μάθησης, η ειδοποιός διαφορά της με την επιβλεπόμενη είναι πως τα δεδομένα εισόδου δεν περιλαμβάνουν τις ετικέτες, είναι δηλαδή unlabelled. Αυτό δεν σημαίνει πως δεν εφαρμόζονται σε δεδομένα για τα οποία



Σχήμα 2.1: Παραδείγματα αλγόριθμων clustering

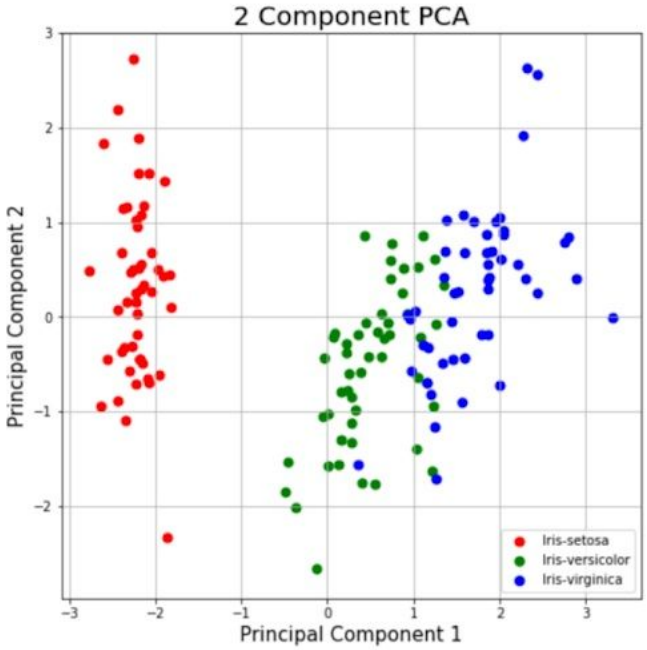
έχουμε τις ετικέτες, αλλά απλώς ότι δεν μπαίνουν σαν είσοδος στα μοντέλα μη επιβλεπόμενης μάθησης. Ο στόχος εδώ είναι να βρεθεί δομή στα δεδομένα ή και να χωριστούν οι παρατηρήσεις σε ομάδες.

Πολλοί διακρίνουν δύο υποκατηγορίες και σε αυτό τον τύπο μάθησης:

- **Συσταδοποίηση (Clustering):** Αφορά τις τεχνικές που χωρίζουν τα δείγματα σε συστάδες, γνωστές και ως clusters. Η κάθε παρατήρηση ανατίθεται σε μία συστάδα (ή ομάδα) και η αυτή ομάδα περιέχει παρατηρήσεις κατά κάποιο τρόπο όμοιες μεταξύ τους. Η ομοιότητα αυτή μπορεί να βασίζεται στην απόσταση των χαρακτηριστικών μεταξύ των παρατηρήσεων.

  Στην εικόνα 2.1 παρουσιάζονται διάφοροι αλγόριθμοι συσταδοποίησης σε δεδομένα διαφόρων σχημάτων. Σε κάθε γραφική απεικόνιση κάθε σημείο έχει το χρώμα που αντιστοιχεί στην συστάδα στην οποία ανοίκει.

- **Μείωση διαστάσεων (Dimensionality reduction)**: Αφορά τις τεχνικές οι οποίες μειώνουν τις διαστάσεις των δεδομένων. Οι τεχνικές αυτές συμπιέζουν τα δεδομένα σε έναν πίνακα μικρότερων διαστάσεων, ο οποίος όμως έχει διατηρήσει την περισσότερη δυνατή πληροφορία από τα αρχικά δεδομένα. Για παράδειγμα, ένας πίνακας δεδομένων με 1000 παρατηρήσεις και 400 χαρακτηριστικά μπορεί να συμπιεσθεί σε μερικές δεκάδες χαρακτηριστικά, καθώς και να προβληθεί στον δυσδιάστατο ή τρισδιάστατο χώρο.



Σχήμα 2.2: Παραδείγματα μείωσης διαστάσεων με PCA

Οι τεχνικές αυτές είναι πολύ χρήσιμες για να οπτικοποιηθούν τεράστιοι όγκοι δεδομένων σε ένα γράφημα δύο ή τριών διαστάσεων. Μια ευρέως χρησιμοποιούμενη τεχνική dimensionality reduction είναι η PCA ή Principal Component Analysis.

- **Ενισχυτική μάθηση (Reinforcement Learning):** Η ενισχυτική μάθηση είναι ένας τύπος μηχανικής μάθησης, όπου ο αλγόριθμος, ή στην συγκεκριμένη περίπτωση πράκτορας, μαθαίνει μέσα από την αλληλεπίδρασή του με το περιβάλλον. Μετά από κάθε κίνηση ο πράκτορας λαμβάνει feedback από το περιβάλλον, ή αλλιώς ανταμοιβή. Κατά αυτόν τον τρόπο, το σύστημα αποκτά γνώση μέσω της εμπειρίας του και έχει ως σκοπό να αναπτύξει την καλύτερη πολιτική (policy) κατά την οποία ελαχιστοποιείται το κόστος και μεγιστοποιείται η απόδοση του συστήματος. Παράδειγμα ενισχυτικής μάθησης θα μπορούσε να είναι ένα ρομπότ που μαθαίνει να πλοηγείται στον χώρο.

2.2.3 Υπολογιστική Όραση

Η υπολογιστική όραση (Computer Vision) είναι ένας τομέας της τεχνητής νοημοσύνης ο οποίος αφορά τεχνικές επεξεργασίας και ανάλυσης εικόνων, καθώς και διάφορους αλγόριθμους εξαγωγής χαρακτηριστικών. Η βασική βλέψη είναι το εκάστοτε υπολογιστικό σύστημα να κατανοήσει την εικόνα μέσω αυτών των χαρακτηριστικών, όπως τα χρώματα της εικόνας, οι γωνίες, ο φωτισμός κ.α.

Η υπολογιστική όραση μπορεί να τελέσει διάφορες διεργασίες. Η πιο συνήθης από αυτές είναι η αναγνώριση αντικειμένων, όπου ένας αλγόριθμος, συνήθως ένα τεχνητό νευρωνικό δίκτυο, εκπαιδεύεται με σκοπό να αναγνωρίσει διάφορα αντικείμενα

Σχήμα 2.3: Παράδειγμα εντοπισμού αντικειμένων με χρήση υπολογιστικής όρασης [2]

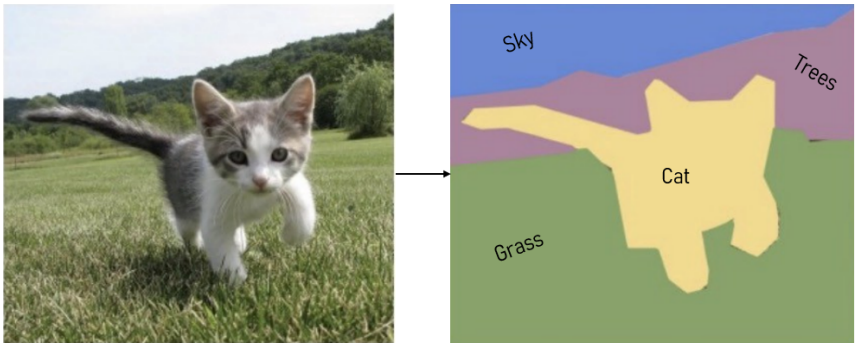
μέσα σε μια εικόνα. Το ίδιο φυσικά μπορεί να γίνει και με πρόσωπα. Αντίστοιχα, ένα μοντέλο θα μπορούσε να ανιχνεύσει ένα αντικείμενο ή ένα πρόσωπο. Η αναγνώριση και η ανίχνευση είναι δύο ελαφρώς διαφορετικές εργασίες, με την πρώτη να ταξινομεί τα αντικείμενα που έχουν βρεθεί (classification), και την δεύτερη απλώς να εντοπίζει την περιοχή στην οποία βρίσκεται εν λόγω αντικείμενο μέσα στην εικόνα (detection).

Μια διαφορετική διεργασία είναι η τμηματοποίηση εικόνων (image segmentation), όπου έχουμε πάλι μια διαδικασία ταξινόμησης. Ωστόσο, σε αυτή την περίπτωση αυτά που ταξινομούνται είναι τα pixels της εικόνας και όχι η ίδια η εικόνα ως σύνολο. Σε κάθε pixel θα πρέπει να ανατεθεί μία και μόνο μία ετικέτα. Η τμηματοποίηση μίας εικόνας έχει ως στόχο τον διαχωρισμό της σε περιοχές που αναλογούν σε ένα και μόνο αντικείμενο.

Τέτοιοι αλγόριθμοι βρίσκουν εφαρμογές σε διάφορους τομείς, όπως οι ιατρικές απεικονίσεις ή η αυτόνομη οδήγηση. Τα αυτοκίνητα που διαθέτουν λειτουργία αυτόματου πιλότου επιστρατεύουν τέτοια μοντέλα για να ξεχωρίσουν διάφορα αντικείμενα από την άσφαλτο.

2.2.4 Άλλοι τομείς

Το πεδίο της τεχνητής νοημοσύνης εξελίσσεται διαρκώς και, όπως είναι αναμενόμενο, έχει αρκετούς τομείς. Ένας άλλος τομέας της είναι η ρομποτική. Η ρομποτική συνδυάζει την τεχνητή νοημοσύνη με την μηχανική με σκοπό την δημιουργία έξυπνων μηχανών, οι οποίες μπορούν να αλληλεπιδράσουν με τον φυσικό κόσμο. Εφαρμογές της ρομποτικής βρίσκει κανείς στην βιομηχανία, στην ιατρική, ακόμα και στο πεδίο της άμυνας.



Σχήμα 2.4: Παράδειγμα τμηματοποίησης εικόνας

Ακόμη, ένας ευρεία γνωστός τομέας είναι η επεξεργασία φυσικής γλώσσας (Natural Language Processing - NLP). Ο τομέας αυτός πραγματεύεται την αλληλεπίδραση των υπολογιστών με την φυσική γλώσσα. Με τις τεχνικές της NLP μπορεί κανείς να αναγνωρίσει και να αναλύσει τα διάφορα δομικά στοιχεία και χαρακτηριστικά ενός κειμένου, καθώς και να προβεί σε πιο δύσκολες διεργασίες όπως είναι η αυτόματη μετάφραση. Τα chatbots κάνουν χρήση τέτοιων τεχνικών για να συνομιλήσουν με τους ανθρώπους με φυσική γλώσσα.

2.3 Αλγόριθμοι μηχανικής μάθησης

2.3.1 Αλγόριθμοι επιβλεπόμενης μάθησης

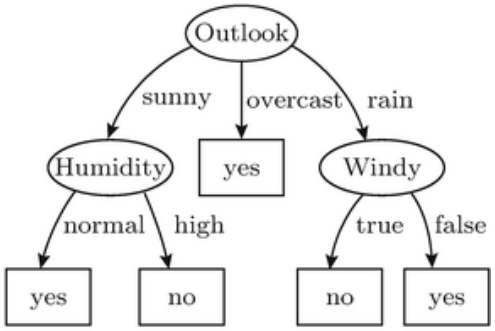
- **Δέντρα αποφάσεων (Desicion Tree)**: Τα δέντρα αποφάσεων χρησιμοποιούνται σε προβλήματα ταξινόμησης αλλά και παλινδρόμησης. Ένα δέντρο απόφασης αναπαριστάται γραφικά ως ένα δέντρο του οποίου κάθε κόμβος αντιστοιχεί σε ένα κριτήριο και κάθε φύλλο σε μία απόφαση ή πρόβλεψη. Ο αλγόριθμος αποφασίζει ποιο κριτήριο θα εφαρμόσει κάθε φορά σε κάθε κόμβο για να χωρίσει τα δεδομένα σε υπο-κατηγορίες. Όταν το δέντρο σχηματιστεί, με βάση όλα τα δεδομένα εκπαίδευσης, τότε μπορούμε να το διατρέξουμε για να πάρουμε μια πρόβλεψη.

  Στην εικόνα 2.5 αναπαριστάται ένα δέντρο απόφασης που αποφασίζει αν κάποιος θα παίξει γκολφ, ανάλογα με την όψη του καιρού, την θερμοκρασία, την υγρασία και τον αέρα [13].

- **K-Nearest Neighbors** Ο αλγόριθμος K-Nearest Neighbors είναι ένας αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται εξίσου για ταξινόμηση και παλινδρόμηση. Η λειτουργία του βασίζεται στην ιδέα ότι δείγματα με παρόμοια χαρακτηριστικά, επομένως δείγματα με μεταξύ τους μικρή απόσταση, τείνουν να έχουν την ίδια ετικέτα, δηλαδή να ανήκουν στην ίδια κλάση.

  Ο ακέραιος αριθμός K αντιπροσωπεύει τον αριθμό των κοντινών στοιχείων, ή αλλιώς γειτόνων, βάσει των οποίων θα γίνει η πρόβλεψη, και ορίζεται από τον χρήστη. Αφού προσδιοριστεί το K, ο αλγόριθμος εκπαιδεύεται απλώς αποθη-

Outlook	*Temp*	*Humidity*	*Windy*	*Golf?*
rainy	hot	high	false	no
rainy	hot	high	true	no
overcast	hot	high	false	yes
sunny	mild	high	false	yes
sunny	cool	normal	false	yes
sunny	cool	normal	true	no
overcast	cool	normal	true	yes
rainy	mild	high	false	no
rainy	cool	normal	false	yes
sunny	mild	normal	false	yes
rainy	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
sunny	mild	high	true	no



Σχήμα 2.5: Παράδειγμα δέντρου απόφασης

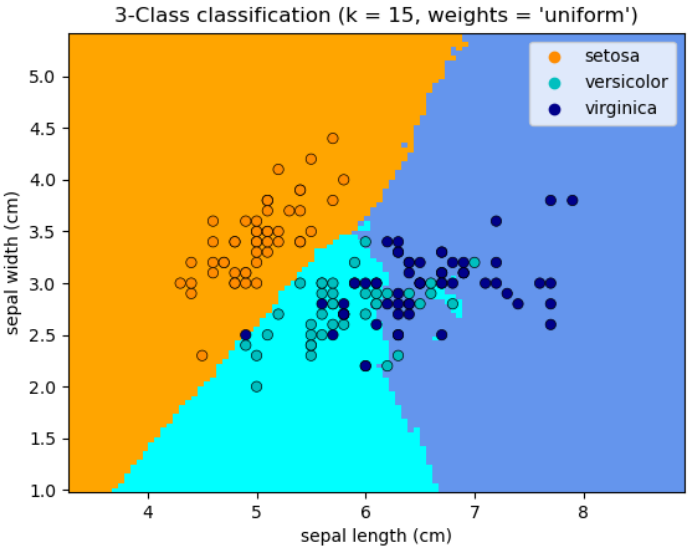
κεύοντας όλα τα δεδομένα εκπαίδευσης στην μνήμη. Έπειτα, όταν έρθει η ώρα πρόβλεψης ενός νέου δείγματος, υπολογίζει την απόσταση ανάμεσα στο νέο δείγμα και σε όλα τα δείγματα εκπαίδευσης, χρησιμοποιώντας κάποια μετρική απόστασης, όπως η Ευκλείδια απόσταση. Εν συνεχεία, επιλέγονται τα Κ κοντινότερα σε απόσταση δείγματα. Τέλος, η κλάση των γειτόνων που εμφανίζεται πιο συχνά θα είναι και η προβλεπόμενη κλάση.

Η πιο συχνή μετρική απόστασης που χρησιμοποιείται είναι η Ευκλείδια απόσταση. Ωστόσο, δεν είναι σπάνιο να χρησιμοποιηθεί κάποια άλλη μέθοδος υπολογισμού απόστασης, ανάλογα με την φύση των δεδομένων. Συχνά συναντάμε την απόσταση Manhattan που υπολογίζει την απόσταση μεταξύ δυο σημείων ως το άθροισμα των απόλυτων διαφορών των συντεταγμένων τους. Ακόμη, υπάρχει η απόσταση Hamming η οποία χρησιμεύει στην εύρεση απόστασης μεταξύ δυο δυαδικών διανυσμάτων, ή η απόσταση cosine που μετράει το συνημίτονο της γωνίας μεταξύ των διανυσμάτων και καθορίζει εάν δύο διανύσματα δείχνουν προς την ίδια κατεύθυνση. Τέλος, υπάρχουν και άλλες μετρικές αποστάσεων όπως η απόσταση Minkowski ή η απόσταση Jaccard.

Παρόλη την ευελιξία και την ευκολία του, το μεγαλύτερο μειονέκτημα του K-NN είναι πως απαιτεί να αποθηκευτούν όλα τα δεδομένα στη μνήμη. Συνεπώς, η απόδοσή του δύναται να επηρεαστεί από τον αριθμό των δεδομένων. Επίσης, σε περιπτώσεις μη ισορροπημένων κλάσεων, μπορεί να μεροληπτεί υπέρ της κλάσης πλειοψηφίας.

Ο αλγόριθμος K-Nearest Neighbors εφαρμόζεται, εκτός άλλων, και σε περιπτώσεις ανίχνευσης ανωμαλιών, όπου εντοπίζει με ευκολία δείγματα που απέχουν σημαντικά από την πλειονότητα των δεδομένων, αλλά και σε συστήματα συστάσεων (Recommender Systems), όπου μπορεί να παράξει συστάσεις προϊόντων με βάση παλαιότερες προτιμήσεις των αγοραστών.

- **Naïve Bayes:** Ο αλγόριθμος Naïve Bayes είναι ένας επίσης απλός και ευρέως



Σχήμα 2.6: Παράδειγμα ορίων απόφασης για ταξινόμηση με K=15

χρησιμοποιούμενος αλγόριθμος ταξινόμησης. Το επίθετο "naïve", δηλαδή αφελής, στο όνομα του οφείλεται στο ότι υποθέτει πως όλα τα χαρακτηριστικά είναι ανεξάρτητα μεταξύ τους, μια υπόθεση που επιβεβαιώνεται σπάνια στην πραγματική ζωή. Εντούτοις, παρόλη την "αφέλειά" του, ο Naïve Bayes έχει επιδείξει πολύ καλή απόδοση σε πολλά πραγματικά προβλήματα ταξινόμησης.

Η λειτουργία του αλγόριθμου βασίζεται στο γνωστό σε όλους Θεώρημα του Bayes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

ο οποίος υπολογίζει την πιθανότητα μίας ετικέτας δεδομένων των χαρακτηριστικών της. Η εκπαίδευσή του γίνεται υπολογίζοντας τις συχνότητες εμφάνισης των χαρακτηριστικών για κάθε κατηγορία. Στην συνέχεια, κατά τη διάρκεια του (testing), αυτές οι πληροφορίες χρησιμοποιούνται για να υπολογιστεί η πιθανότητα κάθε κατηγορίας για μια καινούρια παρατήρηση και για να ταξινομηθεί στην πιο πιθανή κλάση.

Ο Naïve Bayes είναι πρακτικός για προβλήματα ταξινόμησης κειμένου, όπως η αναγνώριση θεμάτων, η αναγνώριση συναισθήματος στο κείμενο ή ο διαχωρισμός ηλεκτρονικής αλληλογραφίας σε spam και μη-spam. Ένα πλεονέκτημά του είναι ότι είναι γρήγορος και δεν χρειάζεται πολλές παραμέτρους.

Υπάρχουν διάφορα είδη αλγορίθμων Naïve Bayes που χρησιμοποιούνται στην μηχανική μάθηση. Μεταξύ τους διαφέρουν στην υπόθεση που κάνουν για την κατανομή των χαρακτηριστικών. Ο πιο γνωστός αλγόριθμος είναι ο Gaussian

Naïve Bayes ο οποίος χρησιμοποιείται όταν θεωρείται πως τα χαρακτηριστικά ακολουθούν κανονική κατανομή όντας μεταξύ τους ανεξάρτητα.

Έστω $P(C_k)$ η πιθανότητα μίας κλάσης, η οποία εκφράζει τον αριθμό των δειγμάτων στην κλάση C_k διαιρεμένο με τον συνολικό αριθμό δειγμάτων. Τότε η πιθανότητα του χαρακτηριστικού θα είναι

$$P(x_i|C_k) = \frac{1}{\sqrt{2\pi\sigma_{C_k}^2}} \exp\left(-\frac{(x_i - \mu_{C_k})^2}{2\sigma_{C_k}^2}\right)$$

όπου μ_{C_k} είναι ο μέσος όρος του χαρακτηριστικού x_i στην κλάση μ_{C_k} και $\sigma_{C_k}^2$ είναι η διακύμανση της τιμής του χαρακτηριστικού x_i στην κλάση μ_{C_k}.

Ένας ακόμη τύπος Naïve Bayes είναι ο Multinomial Naïve Bayes, ο οποίος χρησιμοποιείται συχνά για ταξινόμηση κειμένου με βάση τη συχνότητα εμφάνισης των λέξεων.

- **Γραμμική παλινδρόμηση (Linear Regression)**: Όπως μαρτυρά και το όνομα της, η γραμμική παλινδρόμηση χρησιμοποιείται για την πρόβλεψη μίας συνεχούς μεταβλητής, ή αλλιώς της εξαρτημένης μεταβλητής. Στην γραμμική παλινδρόμηση ο στόχος είναι να προσαρμοστεί μια γραμμική συνάρτηση στα δείγματα εκπαίδευσης, έτσι ώστε να εκτιμηθεί η σχέση μεταξύ των μεταβλητών.

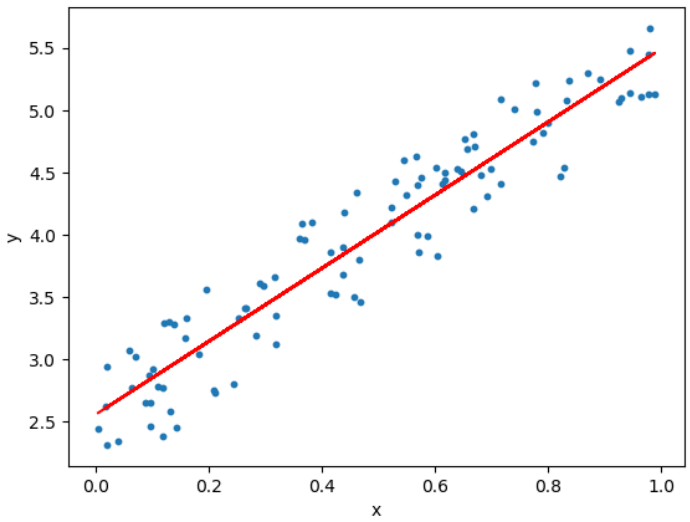
  Ο αλγόριθμος γραμμικής παλινδρόμησης υποθέτει πως οι ανεξάρτητες μεταβλητές μπορούν να προβλέψουν γραμμικά την εξαρτημένη μεταβλητή. Η γενική μορφή της είναι:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$$

  όπου y είναι η εξαρτημένη μεταβλητή, δηλαδή ο στόχος της πρόβλεψης, x_1, x_2, ..., x_n είναι οι ανεξάρτητες μεταβλητές και β_0, β_1, ... , β_n είναι οι παράμετροι της παλινδρόμησης που θα υπολογιστούν κατά τη διάρκεια της εκπαίδευσης του μοντέλου. Σκοπός είναι οι παράμετροι αυτές να πάρουν τις βέλτιστες τιμές για να μειωθεί στο ελάχιστο το σφάλμα μεταξύ προβλέψεων και πραγματικών τιμών.

  Υπάρχουν αρκετές μέθοδοι που εκτιμούν τους συντελεστές αυτούς. Μια από τις πιο δημοφιλείς είναι η Ordinary Least Squares (OLS) η οποία ελαχιστοποιεί το άθροισμα των τετραγώνων των διαφορών μεταξύ των πραγματικών τιμών και των προβλέψεων. Άλλες μέθοδοι περιλαμβάνουν τις Mean Squared Error (MSE) και Mean Absolute Error (MAE) οι οποίες μετράνε το μέσο των τετραγώνων των διαφορών και το μέσο των απόλυτων διαφορών αντίστοιχα.

  Στην εικόνα 2.7 παρουσιάζεται ένα παράδειγμα γραμμικής παλινδρόμησης. Με μπλε χρώμα οπτικοποιούνται τα δείγματα και με κόκκινο χρώμα η ευθεία



Σχήμα 2.7: Παράδειγμα γραμμικής παλινδρόμησης

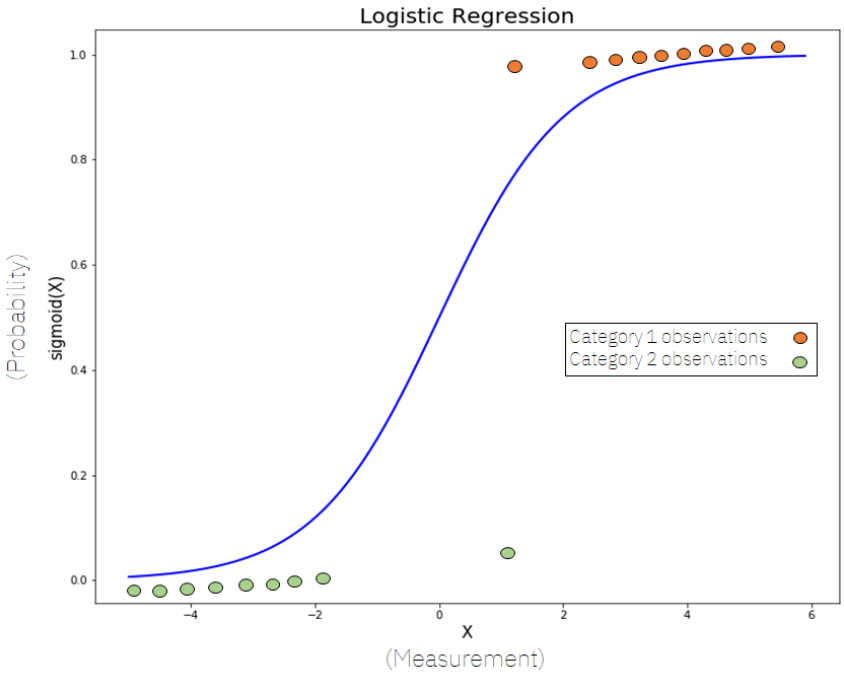
$y = ax + b$ η οποία προσαρμόστηκε στα δεδομένα. Όσο πιο κοντά βρίσκεται ένα σημείο στην ευθεία, τόσο καλύτερη είναι η πρόβλεψή του.

- **Λογιστική παλινδρόμηση (Logistic Regression)**: Η λογιστική παλινδρόμηση είναι ένας αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται αποκλειστικά για ταξινόμηση και κυρίως για δυαδική ή πολυκατηγορική ταξινόμηση, και όχι για παλινδρόμηση. Ονομάζεται έτσι γιατί χρησιμοποιεί την λογιστική συνάρτηση (logistic function) για να τελέσει την ταξινόμηση. Στον συγκεκριμένο αλγόριθμο, η εξαρτημένη μεταβλητή είναι κατηγορική και μπορεί να πάρει δύο ή παραπάνω τιμές. Ο σκοπός είναι η πρόβλεψη της κατηγορίας ενός παραδείγματος βάσει των ανεξάρτητων μεταβλητών. Το όνομα της προκύπτει από την λογιστική συνάρτηση (logistic function), της οποίας οι παράμετροι είναι αυτοί που ρυθμίζονται κατά τη διάρκεια της εκπαίδευσης, και δίνεται από τον τύπο:

$$p = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n)}}$$

Η λογιστική συνάρτηση (βλ. εικόνα 2.8) μετασχηματίζει την είσοδο δίνοντας έξοδο από το 0 μέχρι το 1, εκφράζοντας έτσι την πιθανότητα το δείγμα να ανήκει στην θετική κλάση. Αν η πιθανότητα αυτή είναι μεγαλύτερη από ένα κατώφλι (συνήθως το κατώφλι ορίζεται στο 0.5), τότε το δείγμα ταξινομείται ως "θετικό". Παραδείγματος χάρη, στο κλασικό παράδειγμα ταξινόμησης ηλεκτρονικής αλληλογραφίας ως επιθυμητή ή μη (spam or not spam), ένα e-mail ταξινομείται ως spam αν η πιθανότητα p είναι μεγαλύτερη του 0.5.

- **Τυχαία Δάση (Random Forests)**: Τα τυχαία δάση [14] είναι ένας δημοφιλής



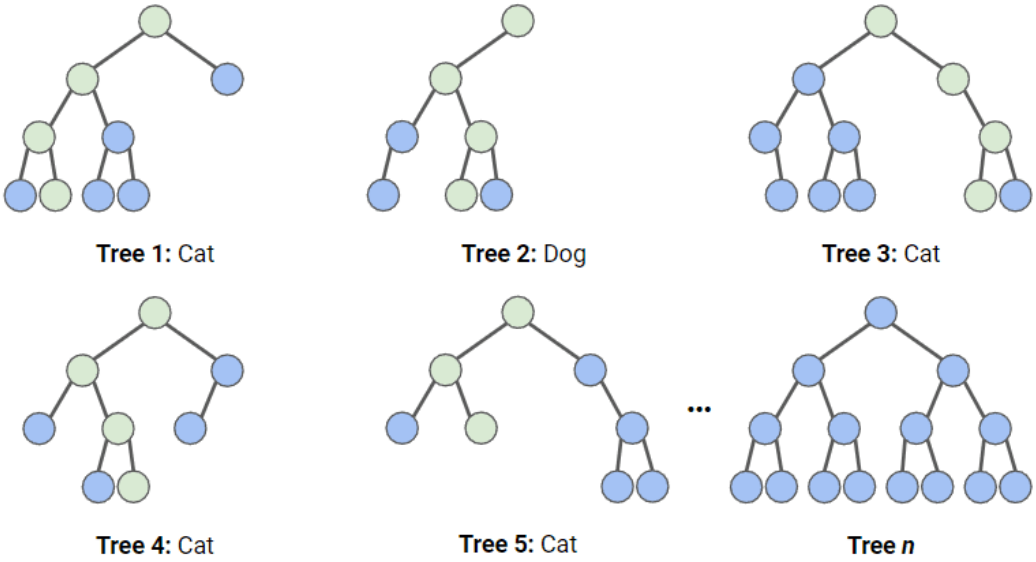
Σχήμα 2.8: Γραφική παράσταση λογιστικής συνάρτησης

αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται για προβλήματα ταξινόμησης αλλά και παλινδρόμησης. Ένα τυχαίο δάσος αποτελείται από πολλά δέντρα αποφάσεων, γνωστά και ως δέντρα αποφάσεων τύπου ensemble.

Κάθε δέντρο απόφασης στο τυχαίο δάσος χτίζεται από μια τυχαία επιλογή κάποιων χαρακτηριστικών. Με αυτόν τον τρόπο δημιουργούνται πολλά δέντρα αποφάσεων που διαφέρουν μεταξύ τους, καθιστώντας έτσι το σύνολο ικανό να αντιμετωπίσει την ποικιλία των δεδομένων.

Κατά τη διάρκεια της πρόβλεψης, το τυχαίο δάσος συνδυάζει τις προβλέψεις κάθε δέντρου για να λάβει μια τελική απόφαση. Στις περιπτώσεις προβλημάτων ταξινόμησης αυτό μπορεί να γίνει παίρνοντας την ψήφο της πλειοψηφίας, δηλαδή η κλάση που προβλέφθηκε από τα περισσότερα δέντρα θα είναι και η πρόβλεψη. Αντίστοιχα, στις περιπτώσεις παλινδρόμησης μπορεί να γίνει παίρνοντας τον μέσο όρο των προβλέψεων όλων των δέντρων. Στην εικόνα 2.9 παρουσιάζεται ένα τυχαίο δάσος εκπαιδευμένο για πρόβλημα ταξινόμησης δύο κλάσεων (σκύλος και γάτα).

Τα τυχαία δάση πλεονεκτούν υπέρ άλλων μοντέλων καθώς συνήθως αποφέρουν υψηλή απόδοση, εφόσον συνδυάζουν την δύναμη πολλών δέντρων αποφάσεων. Επίσης, είναι ανθεκτικά στην υπερεκπαίδευση και το overfitting, αφού κάθε δέντρο δημιουργείται από τυχαίες επιλογές χαρακτηριστικών. Αν ένα από τα δέντρα κάνει μια λάθος πρόβλεψη, είναι πολύ πιθανό ένα άλλο δέντρο να αποζημιώσει κάνοντας την σωστή πρόβλεψη. Ακόμη, τα τυχαία δάση είναι ικανά να αντιμετωπίσουν το φαινόμενο των απόντων τιμών. Τέλος, έχουν την ικανότητα



Σχήμα 2.9: Οπτικοποίηση Τυχαίου Δάσους

να αντιμετωπίσουν διάφορους τύπους δεδομένων και χαρακτηριστικών.

- **Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)**: Οι Μηχανές Διανυσμάτων Υποστήριξης είναι ένας αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται σε προβλήματα παλινδρόμησης και ταξινόμησης. Τα πρώτα SVMs προτάθηκαν από τον Βλαντιμίρ Βάπνικ και τον Αλεξέι Τσερβονένκις το 1964. Τα SVMs βασίζονται στην θεωρία στατιστικής μάθησης.

  Η βασική ιδέα των SVMs είναι η δημιουργία ενός υπερεπιπέδου το οποίο θα διαχωρίζει τις κλάσεις των δεδομένων με τον βέλτιστο τρόπο. Το υπερεπίπεδο αυτό μπορεί να θεωρηθεί ως το σύνορο μεταξύ των δύο κλάσεων και έχει ως στόχο να μεγιστοποιήσει την απόσταση μεταξύ τους και του υπερεπιπέδου (maximum margin), έτσι ώστε να είναι πιο ανθεκτικό σε νέα δεδομένα που δεν έχει ξαναδεί. Ένα τέτοιο υπερεπίπεδο λέγεται υπερεπίπεδο μέγιστου περιθωρίου (maximum margin hyperplane) και για να βρεθεί ο αλγόριθμος βρίσκει δύο ακραία σημεία ή διανύσματα (vectors) που βοηθούν στην δημιουργία του υπερεπιπέδου. Αυτά τα διανύσματα ονομάζονται διανύσματα υποστήριξης και από εκεί ονομάστηκε και ο αλγόριθμος.

  Στην εικόνα 2.10 παρουσιάζονται δύο παραδείγματα οπτικοποίησης μηχανών διανυσμάτων υποστήριξης. Ο στόχος είναι ο διαχωρισμός των τετραγώνων από τους κύκλους. Στο αριστερό μέρος φαίνονται όλα τα μη-βέλτιστα σύνορα που θα μπορούσαν κατά κάποιον τρόπο να διαχωρίσουν τις δύο κλάσεις, ενώ στο δεξί μέρος παρουσιάζεται το υπερεπίπεδο μέγιστου περιθωρίου. Τα τρία σημεία που είναι χρωματισμένα στο κέντρο τους είναι αυτά που οδηγούν την δημιουργία του συνόρου στην μέση και εγγυόνται το μέγιστο περιθώριο μεταξύ των κλάσεων.

Επιπροσθέτως, οι συναρτήσεις πυρήνα (kernel functions) είναι αυτές που επιτρέπουν σε μη-γραμμικά διαχωρίσιμες κλάσεις να διαχωριστούν με την βοήθεια των SVMs. Ουσιαστικά, οι SVMs μετασχηματίζουν τον αρχικό χώρο υποθέσεων για να μετατραπούν σε γραμμικά διαχωρίσιμες και τελικά να διαχωριστούν από το μοντέλο με τον τρόπο που αναφέρθηκε παραπάνω. Οι πιο γνωστές συναρτήσεις πυρήνα είναι η πολυωνυμική (Polynomial kernel), η Γκαουσιανή (Gaussian kernel ή Radial Basis Function - RBF) και η σιγμοειδής (Sigmoid Kernel).

Έστω ένα σύνολο δειγμάτων $x_1, x_2, ..., x_n$ τα οποία ανήκουν σε δύο διαφορετικές κλάσεις, -1 και 1 για την αρνητική και την θετική κλάση αντίστοιχα. Τότε ψάχνουμε το υπερεπίπεδο μέγιστου περιθωρίου έτσι ώστε το κοντινότερο στοιχείο της κάθε κλάσης να έχει την μέγιστη απόσταση από το υπερεπίπεδο αυτό. Το υπερεπίπεδο μπορεί να γραφτεί ως $w_1 \cdot x_1 + w_2 \cdot x_2 + b = 0$ ή $w \cdot x + b = 0$, όπου w τα βάρη που ορίζουν την κατεύθυνση του συνόρου, x το δείγμα και b ένας "όρος μεροληψίας" ή bias term. Ο όρος αυτός επιτρέπει στο σύνορο να αλλάζει κατά μήκος του y άξονα και αντιπροσωπεύει την μετατόπιση του ορίου απόφασης.

Αν α και β δύο σημεία που βρίσκονται πάνω στο σύνορο, εφαρμόζοντας την σχέση $w \cdot x + b = 0$ και αφαιρόντας κατά μέλη, προκύπτει ότι:

$$w(x_\alpha - x_\beta) = 0$$

το οποίο σημαίνει πως το διάνυσμα w είναι κάθετο στο σύνορο.

Σε γραμμικώς διαχωρίσιμα προβλήματα, μπορούν να οριστούν τα δύο επιπλέον διανύσματα εκατέρωθεν του $w \cdot x + b = 0$, που χωρίζουν επίσης τις κλάσεις και το υπερεπίπεδο βρίσκεται στην μέση. Για ένα σημείο z , η κλάση στην οποία ανήκει θα είναι:

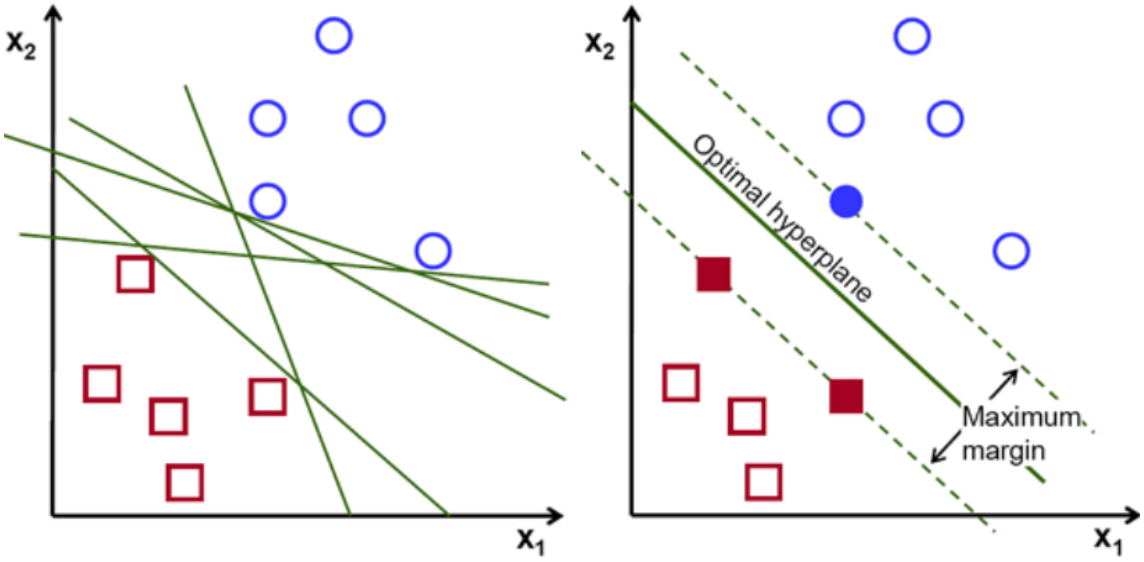
$$y = \begin{cases} 1 & \text{αν } w \cdot z + b > 0 \\ -1 & \text{αν } w \cdot z + b < 0 \end{cases}$$

Αν x_+ και x_- είναι δύο σημεία εκατέρωθεν των ορίων του περιθωρίου, τότε το διάνυσμα w, αφού είναι κάθετο στο σύνορο, αν διαιρεθεί με το μήκος του $||w||$ και πολλαπλασιαστεί με το διάνυσμα της διαφοράς των x_+ και x_-, δίνει την απόσταση d μεταξύ τους. Επομένως:

$$d = (x_+ - x_-) \cdot \frac{w}{||w||}$$

Σύμφωνα με την εξίσωση $w \cdot x + b = 0$, αν αντικαταστήσουμε το $x_+ \cdot w$ με το $1 - b$, το $x_- \cdot w$ με το $1 + b$ και απλοποιήσουμε, προκύπτει ότι:

$$d = \frac{2}{||w||}$$



Σχήμα 2.10: Μηχανές Διανυσμάτων Υποστήριξης

Συμπερασματικά, μεγιστοποιώντας την απόσταση d ελαχιστοποιείται το $\|w\|$.

- **Τεχνικές Boosting και Bagging**: Τόσο το Boosting όσο και το Bagging είναι τενχικές ensemble learning οι οποίες χρησιμοποιούνται για την βελτίωση της απόδοσης των μοντέλων. Και οι δύο τεχνικές συνδυάζουν διάφορα μοντέλα για την λήψη αποφάσεων.

  - **Bagging**: Ο όρος bagging προέρχεται από την σύμπτυξη των λέξεων Bootstrap Aggregating. Το bagging λειτουργεί δημιουργώντας τυχαία υποσύνολα δεδομένων από το αρχικό training set και εκπαιδεύοντας ένα μοντέλο για κάθε τέτοιο τυχαίο υποσύνολο. Εν συνεχεία, τα μοντέλα αυτά συνδυάζονται και η τελική απόφαση λαμβάνεται βάσει της πλειοψηφίας των προβλέψεων όλων των μοντέλων ή βάσει του μέσου όρου τους. Ο στόχος του bagging είναι να μειώσει την διακύμανση των μοντέλων και να προσφέρει σταθερότερες προβλέψεις. Τα τυχαία δέντρα (Random Forests), τα οποία αναφέρθηκαν πιο πάνω, χρησιμοποιούν την τεχνική bagging για να δημιουργήσουν όλα τα ξεχωριστά δέντρα που τα αποτελούν.

  - **Boosting**: Το Boosting επικεντρώνεται στην βελτίωση των μοντέλων από άποψη ακρίβειας. Αρχικά, ένα "αδύναμο" μοντέλο εκπαιδεύεται στα αρχικά δεδομένα και οι προβλέψεις του αξιολογούνται. Στη συνέχεια, επικεντρώνεται στα δείγματα που προβλέφθηκαν λάθος και επανεκπαιδεύει ένα νέο μοντέλο που εστιάζει μόνο σε αυτά τα δείγματα. Αυτή η διαδικασία επαναλαμβάνεται πολλές φορές με την δημιουργία πολλών μοντέλων που συνεχώς βελτιώνονται. Το τελικό μοντέλο που θα προκύψει θα είναι ο συνδυασμός των εκπαιδευμένων μοντέλων. Το boosting εχει ως στόχο να δημιουργήσει ένα ισχυρό μοντέλο υψηλής απόδοσης.

Ένας δημοφιλής αλγόριθμος μηχανικής μάθησης ο οποίος χρησιμοποιέι boosting είναι ο Gradient Boosting, ο οποίος εκπαιδεύει μοντέλα, συνήθως δέντρα αποφάσεων, χρησιμοποιώντας τον αλγόριθμο gradient descent. Ο σκοπός εδώ είναι να ελαχιστοποιηθεί η συνάρτηση κόστους. Ένας άλλος αλγόριθμος είναι ο AdaBoost, ο οποίος χρησιμοποιεί κατά κανόνα decision stumps (δηλαδή μικρά δέντρα αποφάσεων με μία μόνο διαίρεση). Ο AdaBoost επίσης εκπαιδεύει σειριακά ένα ensemble από αδύναμα μοντέλα, δίνοντας έμφαση στα παραδείγματα που δεν προβλέφθηκαν σωστά, και στο τέλος συνδυάζει τις προβλέψεις όλων των αδύναμων μοντέλων εκχωρώντας μεγαλύτερο βάρος στις περιπτώσεις που δεν προβλέψθηκαν σωστά.

Τέλος, ένας ακόμη αλγόριθμος που αξιοποιεί την τεχνική του boosting είναι ο XGBoost. Ο XGBoost αποτελεί εξέλιξη του Gradient Boosting και χρησιμοποιεί έναν πολύπλοκο συνδυασμό αδύναμων μοντέλων, εκπαιδεύοντάς τα με παράλληλο τρόπο για αυξημένη ταχύτητα και απόδοση. Τέλος, ο XGBoost χρησιμοποιεί λεπτομερείς τεχνικές ρύθμισης για να αποτρέψει το φαινόμενο του overfitting και να βελτιώσει την απόδοση του μοντέλου.

2.3.2 Αλγόριθμοι μη επιβλεπόμενης μάθησης

Όπως αναφέρθηκε και στο υποκεφάλαιο 2.2.1, οι αλγόριθμοι μη επιβλεπόμενης μηχανικής μάθησης δεν δέχονται ως είσοδο τις ετικέτες των δειγμάτων, και χρησιμοποιούνται για να ανακαλύψουν μοτίβα, δομές και συστάδες μέσα στα δεδομένα.

- **Αλγόριθμοι συσταδοποίησης (Clustering Algorithms):** Η συσταδοποίηση είναι μια τεχνική ανάλυσης δεδομένων κατά την οποία τα δείγματα ομαδοποιούνται σε συστάδες (clusters) με βάση την ομοιότητά τους. Οι αλγόριθμοι συσταδοποίησης εξετάζουν την ομοιότητα μεταξύ των δεδομένων με βάση κάποιο μέτρο απόστασης ή ομοιότητας.

  - **K-Means:** Πρόκειται για έναν από τους πιο δημοφιλείς αλγόριθμους συσταδοποίησης. Ο στόχος του είναι να ομαδοποιήσει τα δείγματα σε k συστάδες, όπου k ένας προκαθορισμένος αριθμός συστάδων. Αρχικά, ο αλγόριθμος επιλέγει τυχαία k κέντρα συστάδων από τα δείγματα και ύστερα κάθε δείγμα ανατίθεται σε μια συστάδα, ανάλογα με την απόσταση από το πλησιέστερο κέντρο. Συνήθως χρησιμοποιείται η Ευκλείδια απόσταση. Έπειτα, για κάθε συστάδα, υπολογίζεται το νέο κέντρο της, με βάση τον μέσο όρο των δειγμάτων που της ανήκουν. Τέλος, η διαδικασία αυτή επαναλαμβάνεται μέχρις να σταθεροποιηθούν τα κέντρα των συστάδων.

    Η επιλογή του αριθμού k και η αρχικοποίηση των κέντρων των συστάδων μπορούν να επηρεάσουν τα αποτελέσματα της συσταδοποίησης.

  - **DBSCAN:** Ο αλγόριθμος DBSCAN ή Density-Based Spatial Clustering of

Applications with Noise, είναι ένας αλγόριθμος συσταδοποίησης ο οποίος βασίζεται στην πυκνότητα των δεδομένων. Απαιτεί δύο εισόδους από τον χρήστη, την παράμετρο ϵ και την παράμετρο minPoints. Η πρώτη παράμετρος αναφέρεται στην ακτίνα του κύκλου που θα δημιουργηθεί γύρω από κάθε σημείο δεδομένων για να ελεγχθεί η πυκνότητα, ενώ η δεύτερη αναφέρεται στον ελάχιστο αριθμό σημείων που θα πρέπει να υπάρχουν μέσα σε αυτόν τον κύκλο, έτσι ώστε το σημείο αυτό να μην ταξινομηθεί ως "θόρυβος", δηλαδή να μην ανήκει σε καμία συστάδα.

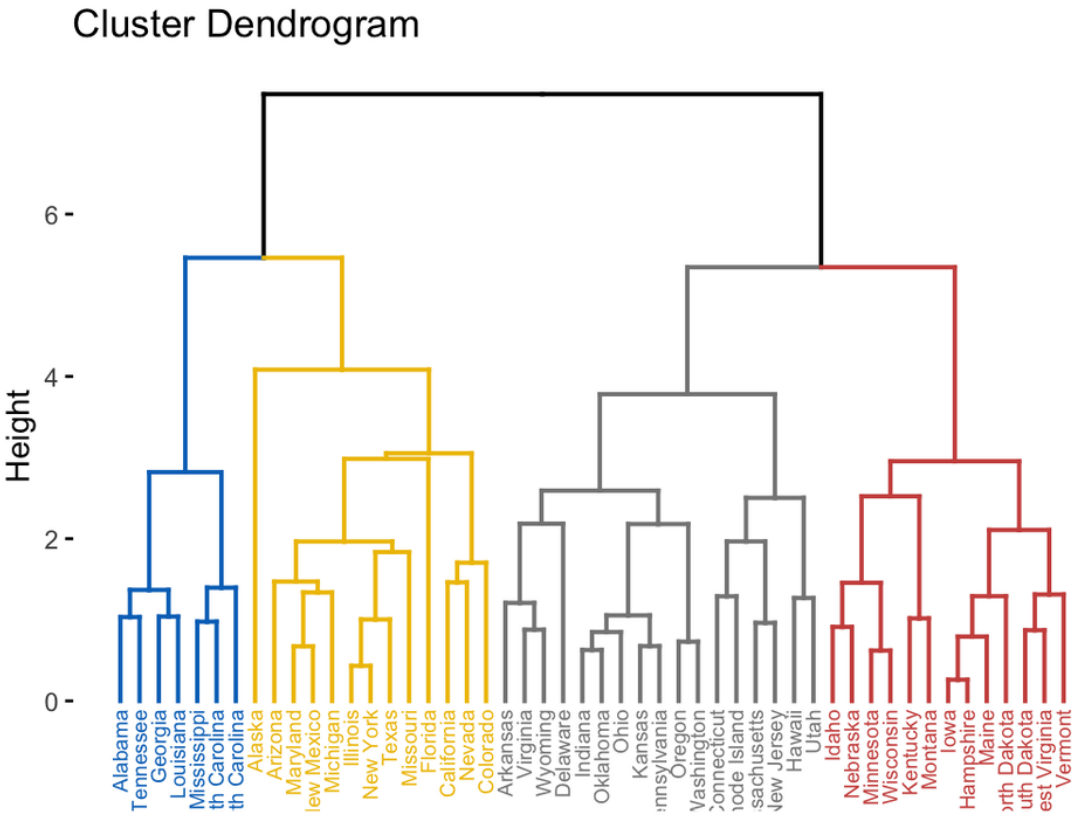
Αρχικά, επιλέγεται ένα τυχαίο σημείο από τα δείγματα. Έπειτα, βρίσκονται όλα τα σημεία σε ακτίνα ϵ από αυτό και τελείται ο έλεγχος πυκνότητας. Αν ο αριθμός δειγμάτων εντός του κύκλου είναι μεγαλύτερος από το προκαθορισμένο minPoints τότε το σημείο θεωρείται πυκνό. Σε αυτή την περίπτωση, ανήκει στην συστάδα με τα γειτονικά του σημεία. Τέλος, αυτή η διαδικασία επαναλαμβάνεται για κάθε σημείο. Ο DBSCAN αποτελεί μια αποτελεσματική μέθοδο συσταδοποίησης, αφού μπορεί να βρει συστάδες σε διάφορα μεγέθη και είναι ανθεκτικός σε ακραίες τιμές. Επίσης, δεν χρειάζεται να προκαθοριστεί ο αριθμός συστάδων, αφού τον βρίσκει αυτόματα.

- **Ιεραρχική συσταδοποίηση (Hierarchical Clustering):** Ο ιεραρχικός αλγόριθμος συσταδοποίησης που χρησιμεύει στην συσταδοποίηση δεδομένων, μειώνοντας την απόσταση μεταξύ των δεδομένων κάθε συστάδας. Ο αλγόριθμος δημιουργεί ένα ιεραρχικό δέντρο συστάδων, γνωστό και ως δεντρόγραμμα (βλ. εικόνα 2.11), που απεικονίζει την ομοιότητα των δεδομένων.

  Υπάρχουν δύο τύποι ιεραρχικής συσταδοποίησης, η συσσωρευτική συσταδοποίηση (agglomerative clustering) και η διχαστική συσταδοποίηση (divisive clustering). Η συσσωρευτική συσταδοποίηση (agglomerative clustering) ξεκινάει θεωρώντας το κάθε δείγμα ως μια ξεχωριστή συστάδα. Έπειτα, αναγνωρίζει δύο συστάδες οι οποίες είναι πιο κοντά και τις ενώνει, επαναλαμβάνοντας την διαδικασία αυτή μέχρι όλες οι συστάδες να ενωθούν.

  Από την άλλη, η διχαστική συσταδοποίηση (divisive clustering) ξεκινάει από μία μεγάλη συστάδα που περιέχει όλα τα δεδομένα και σε κάθε βήμα χωρίζεται σε δύο υπο-συστάδες μέχρι κάθε δείγμα να ανήκει σε μία μοναδική συστάδα. Κατά τη διάρκεια όλης αυτής της διαδικασίας χρησιμοποιούνται διάφορες μετρικές απόστασης για την αξιολόγηση της ομοιότητας μεταξύ των δεδομένων, όπως η Ευκλείδια απόσταση ή η απόσταση Manhattan.

  Η ιεραρχική συσταδοποίηση φέρει αρκετά προτερήματα, όπως το ότι μπορεί να διαχειριστεί μη κυρτές συστάδες καθώς και συστάδες σε διαφορετικά μεγέθη και πυκνότητες. Επίσης, μπορεί να διαχειριστεί δεδομένα



Σχήμα 2.11: Δεντρόγραμμα ιεραρχικής συσταδοποίησης (παράδειγμα)

με θόρυβο. Τέλος, μπορεί να ανακαλύψει την ιεραρχία των δεδομένων, πράγμα χρήσιμο για να κατανοηθεί η σχέση μεταξύ των συστάδων.

- **Αλγόριθμοι μείωσης διαστάσεων:**

Ο χειρισμός δεδομένων υψηλών διαστάσεων αποτελεί μια πρόκληση στο πεδίο της μηχανικής μάθησης. Εαν η διάσταση των δεδομένων αυξηθεί προσθέτοντας επιπλέον χαρακτηριστικά, τότε το μοντέλο μηχανικής μάθησης θα γίνει πιο περίπλοκο. Το πρόβλημα αυτό συχνά αποκαλείται "η κατάρα της διάστασης" ή αλλιώς "the curse of dimensionality". Ως εκ τούτου, συχνά προκύπτει η ανάγκη να μειωθεί ο αριθμός των χαρακτηριστικών, κάτι που δύναται να γίνει μέσω των τεχνικών μείωσης διαστάσεων.

  - **Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis):** Πρόκειται για μια τεχνική μείωσης διαστάσεων η οποία χρησιμοποιείται γαι να εξάγει βασικές πληροφορές από πολυδιάστατα δεδομένα. Η Ανάλυση Κύριων Συνιστωσών έχει σκοπό τον μετασχηματισμό των δεδομένων αυτών σε έναν χώρο χαμηλότερων διαστάσεων, διατηρώντας παράλληλα όσο το δυνατόν περισσότερη πληροφορία. Εν ολίγοις, η PCA συμπιέζει τα δεδομένα προσέχοντας να μην χαθεί πολύτιμη πληροφορία, διατηρώντας την μέγιστη δυνατή διακύμανση των αρχικών δεδομένων.

Η PCA λειτουργεί με την εύρεση των κύριων συνιστωσών των δεδομένων, που είναι γραμμικοί συνδυασμοί των αρχικών μεταβλητών. Αρχικά, κανονικοποιεί καθένα από τα χαρακτηριστικά των δεδομένων, έτσι ώστε να συνεισφέρουν εξίσου στην ανάλυση. Η κανονικοποίηση για κάθε μεταβλητή επιτυγχάνεται αφαιρώντας τον μέσο όρο από κάθε τιμή και διαιρώντας με την τυπική απόκλιση.

$$z = \frac{x - \mu}{\sigma}$$

Εν συνεχεία, υπολογίζει τον πίνακα συνδιακύμανσης (covariance matrix) για τις μεταβλητές, για να αναγνωριστούν οι μεταβλητές που ίσως συσχετίζονται μεταξύ τους σε μεγάλο βαθμό, προσφέροντας ουσιαστικά περιττή πληροφορία. Ο πίνακας συνδιακύμανσης είναι ένας συμμετρικός πίνακας διαστάσεων $p \times p$, όπου p ο αριθμός των διαστάσεων. Ο πίνακας συνδιακύμανσης λαμβάνει υπόψιν όλες τις μεταβλητές και βρίσκει την συνδιακύμανση για όλα τα πιθανά ζευγάρια.

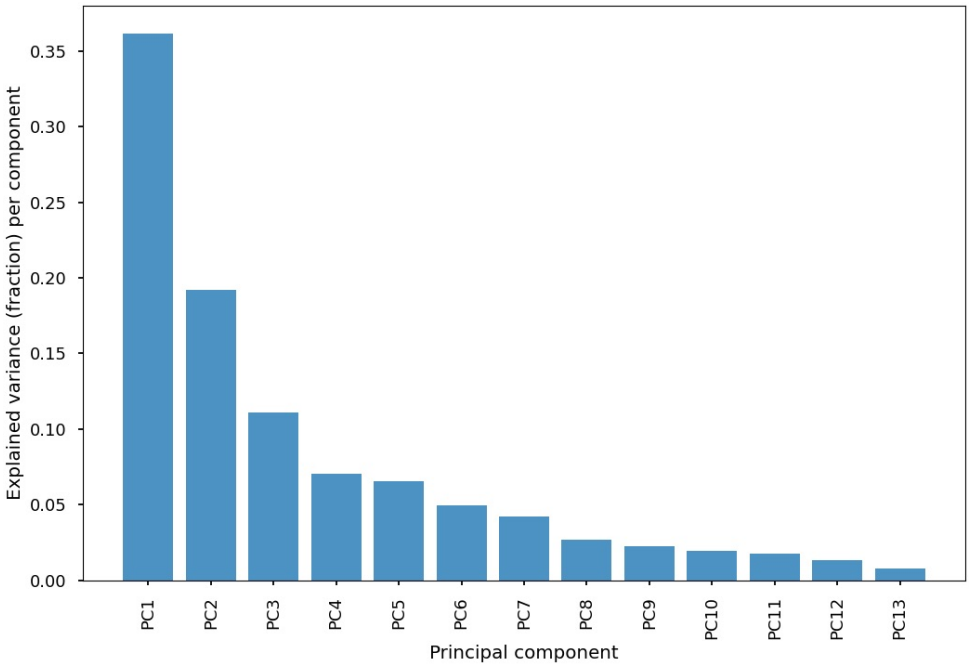
Για παράδειγμα, ο πίνακας συνδιακύμανσης για τρεις μεταβλητές x, y και z θα είναι της μορφής:

$$\text{Covariance Matrix} = \begin{bmatrix} \text{cov}(x,x) & \text{cov}(x,y) & \text{cov}(x,z) \\ \text{cov}(y,x) & \text{cov}(y,y) & \text{cov}(y,z) \\ \text{cov}(z,x) & \text{cov}(z,y) & \text{cov}(z,z) \end{bmatrix}$$

Έπειτα, υπολογίζονται οι ιδιοτιμές και τα ιδιοδιανύσματα του πίνακα διακύμανσης, και ταξινομούνται κατά φθίνουσα σειρά. Με αυτόν τον τρόπο είναι ταξινομημένα με βάση το πόση διακύμανση των αρχικών δεδομένων εξηγούν. Έτσι, η πρώτη κύρια συνιστώσα είναι αυτή που εξηγεί την μεγαλύτερη διακύμανση, δεύτερη την αμέσως επόμενη και ούτω καθεξής. Όπως φαίνεται στην εικόνα 2.12, το ποστοσό της διακύμανσης που εξηγείται από κάθε συνιστώσα (component) μπορεί να οπτικοποιηθει με την βοήθεια ενός ραβδογράμματος. Λαμβάνοντας υπόψιν όλες τις συνιστώσες δύναται να ανακατασκευάσουμε τα δεδομένα με πολύ μεγάλη ακρίβεια. Τέλος, τα δεδομένα προβάλονται στον νέο χώρο που δημιουργήθηκε από τις επιλεγμένες κύριες συνιστώσες.

Ένα βασικό προτέρημα της PCA είναι πως, μετά τον μετασχηματισμό, τα δεδομένα μπορούν να προβληθούν τον χώρο, σε δύο ή και τρεις άξονες, και έτσι να κατανοηθεί και να παρατηρηθεί η δομή τους και το αν συμβαίνει κάποιος τυχόν διαχωρισμός.

– **t-Distributed Stochastic Neighbor Embedding (t-SNE):** Πρόκειται για άλλη μια μέθοδο μείωσης διαστάσεων. Χρησιμοποιείται συνήθως για την οπτικοποίηση δεδομένων υψηλής διάστασης σε έναν χώρο χαμηλότερης



Σχήμα 2.12: Ραβδόγραμμα διακύμανσης που εξηγείται για κάθε συνιστώσα

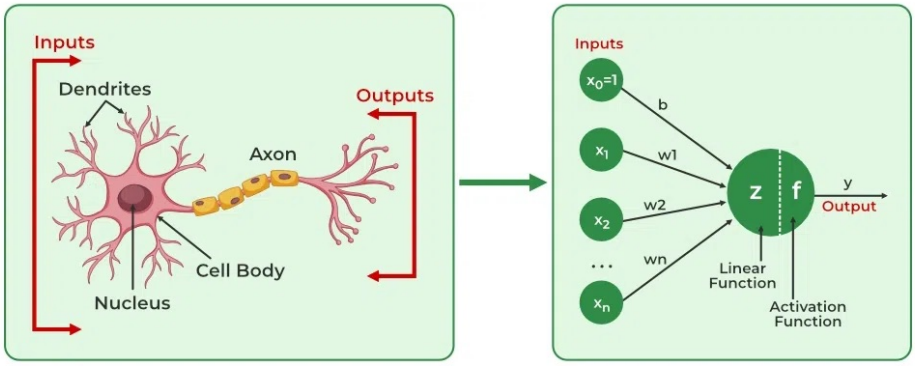
διάστασης και προτάθηκε το 2008 από τους Laurens van der Maaten και Geoffrey Hinton [15]. Αρχικά, το t-SNE υπολογίζει την πιθανότητα ενός ζεύγους σημείων να γειτνιάζουν με βάση τις αποστάσεις τους. Αυτός ο υπολογισμός εκτελείται για κάθε ζεύγος σημείων χρησιμοποιώντας την κατανομή t-Student και έπειτα δημιουργείται ο χώρος χαμηλότερων διαστάσεων. Στον νέο αυτό χώρο, κάθε σημείο αντιπροσωπεύει ένα δείγμα και οι αποστάσεις κάθε σημείου με τα υπόλοιπα υποδηλώνουν την πιθανότητα να είναι γείτονες στον αρχικό χώρο μεγαλύτερων διαστάσεων.

- **Άλλες μέθοδοι μείωσης διαστάσεων:** Φυσικά υπάρχουν και άλλες μέθοδοι μείωσης διαστάσεων, των οποίων οι λεπτομέρειες υπερβαίνουν το πλαίσιο της παρούσας διπλωματικής εργασίας. Επιγραμματικά θα αναφερθούν οι πιο γνωστές μέθοδοι όπως η τεχνική UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) [16], η τεχνική ανάλυσης παραγόντων (Factor Analysis) και η ανάλυση γραμμικών διακρίσεων (Linear Discriminant Analysis - LDA). Συγκεκριμένα, η τελευταία, χρησιμοποιεί την γραμμική διακριτική ανάλυση του Fisher [17] και θεωρείται εξαίρεση καθώς είναι επιβλεπόμενη τεχνική μείωσης διαστάσεων, και όχι μη επιβλεπόμενη.

2.4 Τεχνητά Νευρωνικά Δίκτυα

2.4.1 Τεχνητοί και βιολογικοί νευρώνες

Τα τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks - ANNs) είναι υπολογιστικά μοντέλα τα οποία εμπνεύστηκαν από βιολογικά μοντέλα, δηλαδή από τον τρόπο



Σχήμα 2.13: Αριστερά: Βιολογικός Νευρώνας, δεξιά: τεχνητός νευρώνας

που λειτουργεί ο ανθρώπινος εγκέφαλος. Τα τεχνητά νευρωνικά δίκτυα προσπαθούν ουσιαστικά να μιμηθούν τον τρόπο που συμπεριφέρονται οι νευρώνες του ανθρώπινου εγκεφάλου. Υπάρχουν σε διάφορες μορφές και βρίσκουν εφαρμογές σε μια πληθώρα προβλημάτων όπως η μηχανική μάθηση, η αναγνώριση προτύπων, η αναγνώριση φωνής και η αναγνώριση προσώπων.

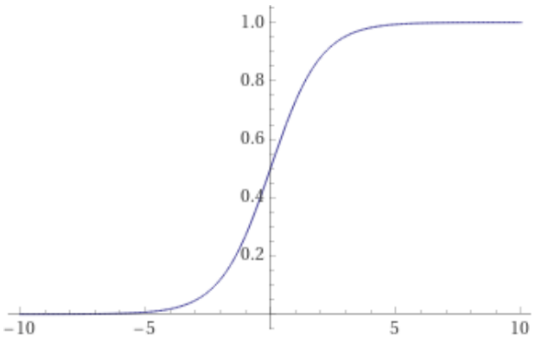
Κάθε νευρώνας αποτελείται από τους δενδρίτες, που λειτουργούν ως είσοδος, το κυρίως σώμα και τον νευροάξονα που συνδέει ένα νευρώνα με άλλους νευρώνες. Κάθε νευρώνας μεταφέρει σήμα από τον άξονά του στον δενδρίτη του άλλου νευρώνα. Το σημείο που ενώνονται αυτά τα δύο ονομάζεται σύναψη ή νευροαξονική απόληξη. Από την άλλη, στα τεχνητά νευρωνικά δίκτυα, κάθε τεχνητός νευρώνας αποτελείται από πολλαπλές εισόδους x_i και μία έξοδο y. Στην εικόνα 2.13 οπτικοποιείται η αντιστοιχία μεταξύ ενός τεχνητού και ενός βιολογικού νευρώνα.

Ο νευρώνας αυτός ονομάζεται και Single Layer Perceptron ή Perceptron του Rosenblatt [18]. Αν και η έννοια του Perceptron επινοήθηκε νωρίτερα από τον Warren McCulloch και τον Walter Pitts (McCulloch-Pitts neuron) [19], ο Franck Rosenblatt πρότεινε μια βελτιωμένη έκδοσή του το 1957.

Το perceptron λαμβάνει πολλές εισόδους $x_1, x_2, ...x_n$, καθώς και έναν όρο "πόλωσης" (bias term) x_0. Καθεμία από τις εισόδους πολλαπλασιάζεται με ένα βάρος w_i, ενώ η πόλωση πολλαπλαζεται με ένα βάρος $w_0 = -\theta$ και αμέσως αθροίζονται. Στο τέλος το άθροισμα αυτό περνάει από μια συνάρτηση ενεργοποίησης ϕ και έτσι λαμβάνουμε την τελική έξοδο y.

$$y_k = \phi(\sum_{i=1}^{n} x_i w_i + b_k)$$

όπου b_k η τελική πόλωση.



Σχήμα 2.14: Γραφική αναπαράσταση σιγμοειδούς συνάρτησης

2.4.2 Συναρτήσεις ενεργοποίησης (Activation functions)

Οι συναρτήσεις ενεργοποίησης είναι αναπόσπαστο κομμάτι των τεχνητών νευρωνικών δικτύων, καθώς αποτελούν την μαθηματική περιγραφή του τρόπου με τον οποίο οι νευρώνες ανταποκρίνονται στο συνολικό ερέθισμα που λαμβάνουν και παράγουν την έξοδο τους. Οι συναρτήσεις ενεργοποίησης είναι υπεύθυνες για την εισαγωγή μη γραμμικότητας (nonlinearity) στα τεχνητά νευρωνικά δίκτυα και επιτρέπουν να μοντελοποιηθούν πιο περίπλοκες συμπεριφορές.

Μερικές απο τις πιο γνωστές συναρτήσεις ενεργοποίησης αποτελούν οι συναρτήσεις που αναφέρονται παρακάτω.

- **Σιγμοειδής (Sigmoid):** Πρόκειται για μια μη γραμμική συνάρτηση η οποία παράγει ως έξοδο έναν αριθμό στο διάστημα [0,1]. Το όνομα της προέρχεται από την μορφή της γραφικής της παράστασης (εικόνα 2.14), η οποία μοιάζει με το αγγλικό γράμμα "S". Δίνεται από την εξίσωση

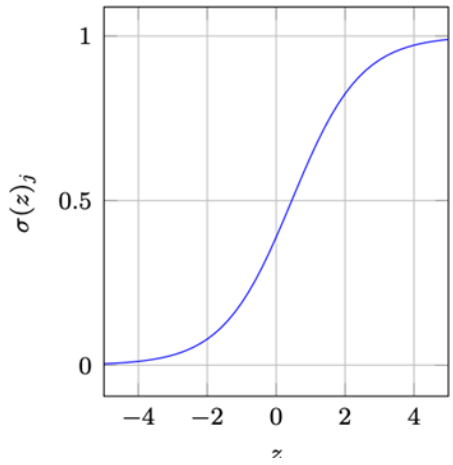
$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$$

  και χρησιμοποιείται συχνά σε προβλήματα ταξινόμησης όπου η έξοδος πρέπει να ερμηνευθεί ως πιθανότητα.

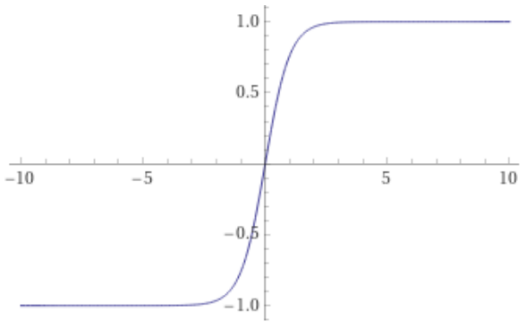
- **Συνάρτηση Softmax:** Η συνάρτηση ενεργοποίησης Softmax χρησιμοποιείται συνήθως στο τελευταίο στρώμα ενός τεχνητού νευρωνικού δικτύου, ιδίως σε προβλήματα ταξινόμησης πολλαπλών ετικετών (multi-class classification), με σκοπό να παράγει την πιθανότητα κάθε κλάσης.

$$\sigma(x)_i = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}}, \quad \text{για } i = 1, 2, \ldots, K$$

  όπου K είναι ο αριθμός των κλάσεων. Η softmax μετατρέπει την είσοδό της σε ένα διάνυσμα με στοιχεία στο [0,1] τα οποία αθροίζουν στο 1. Έτσι, κάθε στοιχείο του διανύσματος αυτού δείχνει την πιθανότητα της εισόδου να ανήκει στην κλάση i.



Σχήμα 2.15: Γραφική αναπαράσταση συνάρτησης softmax



Σχήμα 2.16: Γραφική αναπαράσταση συνάρτησης TanH

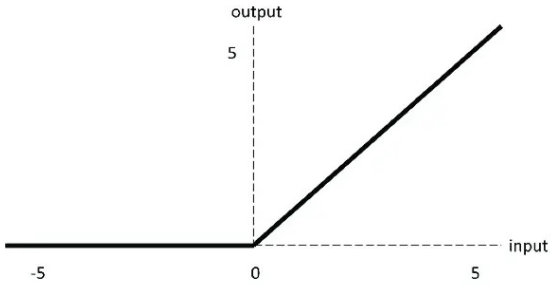
- **Συνάρτηση ενεργοποίησης υπερβολικής εφαπτομένης (TanH):** Είναι μια συνάρτηση ενεργοποίησης παρόμοια με την σιγμοειδή, ωστόσο παράγει έξοδο στο διάστημα [-1, 1]. Η μορφή της είναι επίσης σιγμοειδής αλλά η κλιση της αλλάζει στα όρια του εύρους. Η συνάρτηση αυτή είναι χρήσιμη σε προβλήματα που απαιτούν ενεργοποίηση η οποία να καλύπτει ένα μεγαλύτερο εύρος τιμών.

$$f(x) = \sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

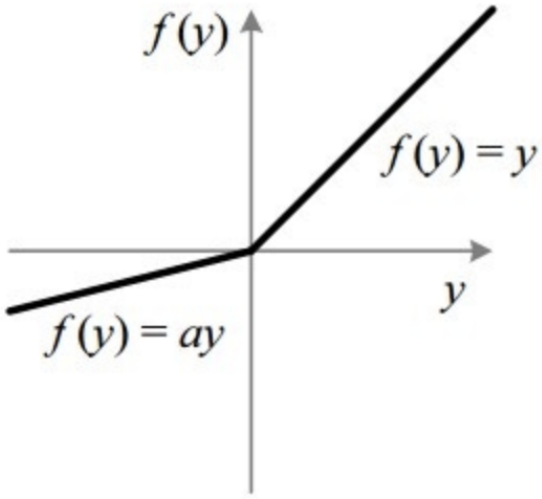
- **Συνάρτηση Ενεργοποίησης ReLU (Rectified Linear Unit):** Η συνάρτηση ReLU είναι μια απλή και αποτελεσματική συνάρτηση η οποία εφαρμόζει γραμμική ενεργοποίηση για θετικές εισόδους και μηδενική ενεργοποίηση για αρνητικές εισόδους. Η συνάρτηση ReLU έχει ως κύριο προτέρημα πως είναι γρήγορη και δίνεται από την ακόλουθη εξίσωση:

$$f(x) = max(0, x)$$

- **Συνάρτηση Ενεργοποίησης Leaky ReLU (Leaky Rectified Linear Unit):**



Σχήμα 2.17: Γραφική αναπαράσταση ReLU



Σχήμα 2.18: Γραφική αναπαράσταση Leaky ReLU

Πρόκειται για μια παραλλαγή της συνάρτησης ενεργοποίησης ReLU, μόνο που αντί για έξοδο 0 στις αρνητικές εισόδους δίνει την τιμή εισόδου πολλαλασιασμένη με μια τιμή a. Προσθέτει, δηλαδή, μια ελαφριά κλίση στην πλερά των αρνητικών τιμών. Δίνεται από την εξίσωση:

$$f(x) = \begin{cases} x, & \text{αν } x \geq 0 \\ ax, & \text{για κάθε άλλη περίπτωση} \end{cases}$$

2.4.3 Perceptron πολλαπλών επιπέδων (Multilayer Perceptron - MLP)

Το Perceptron πολλαπλών επιπέδων (Multilayer Perceptron) είναι ένα είδος τεχνητού νευρωνικού δικτύου το οποίο αποτελείται από πολλαπλά επίπεδα ή στρώματα perceptrons. Είναι το πιο κοινό και δημοφιλές μοντέλο νευρωνικών δικτύων και χρησιμοποιείται εξίσου σε προβλήματα ταξινόμησης αλλά και παλινδρόμησης.

Το MLP αποτελείται από τρία βασικά στοιχεία: το στρώμα εισόδου input layer, ένα ή παραπάνω κρυφά στρώματα (hidden layers) και το στρώμα εξόδου (output layer)

(εικόνα 2.20). Κάθε επίπεδο αποτελείται από πολλούς νευρώνες ή κόμβους (δηλαδή απλά perceptrons) και οι νευρώνες κάθε στρώματος συνδέονται με τους νευρώνες του επόμενου και του προηγούμενου στρώματος. Κάθε σύνδεση έχει και το δικό της βάρος w. Συνήθως κάθε κόμβος ενός στρώματος συνδέεται με όλους τους κόμβους του επόμενου και σε αυτή την περίπτωση έχουμε ένα πλήρως συνδεδεμένο δίκτυο (fully connected), ενώ σε άλλες περιπτώσεις μπορεί να είναι μερικώς συνδεδεμένο (partially connected).

Κατά την διαδικασία εκπαίδευσης, το MLP δέχεται τα δεδομένα ως είσοδο και προσπαθεί να προσαρμόσει ανάλογα τα βάρη των νευρώνων για να μειώσει το σφάλμα της πρόβλεψης. Η διαδικασία αυτή λαμβάνει χώρα με έναν αλγόριθμο που ονομάζεται οπισθοδιάδοση ή κοινώς backpropagation, και επαναλαμβάνεται αρκετές φορές μέχρι ή μέχρι να επιτευχθεί σύγκλιση. Ο αριθμός αυτών των επαναλήψεων ονομάζεται εποχές (epochs).

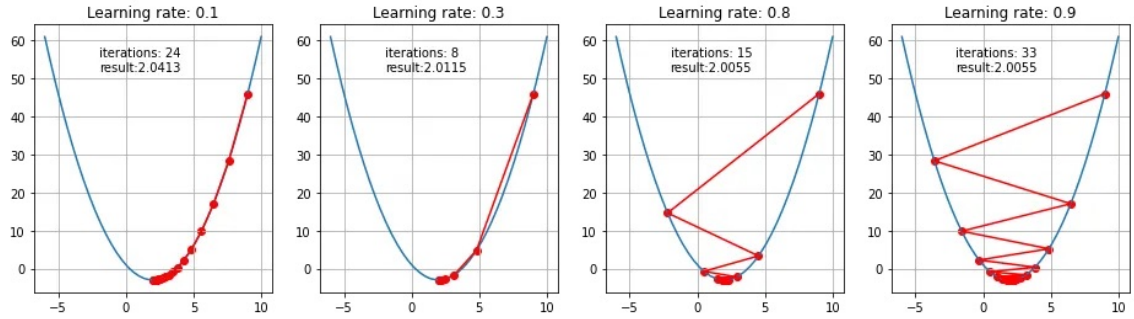
2.4.4 Αλγόριθμος οπισθοδιάδοσης (Backpropagation)

Ο αλγόριθμος οπισθοδιάδοσης, ο οποίος χρησιμοποιείται κατά τη διάρκεια εκπαίδευσης ενός ΤΝΔ, είναι μια μέθοδος η οποία ανανεώνει τα βάρη του ΤΝΔ έχοντας ως στόχο την ελαχιστοποίηση της συνάρτησης κόστους για να παράγει ακριβή αποτελέσματα. Ο αλγόριθμος οπισθοδιάδοσης βασίζεται στον κανόνα της αλυσίδας (chain rule) της θεωρίας παραγώγων. Η βασική ιδέα είνα να υπολογιστούν οι παράγωγοι του κόστους ως προς τα βάρη του δικτύου ξεκινώντας απο το στρώμα εξόδου και με κατεύθυνση προς τα πίσω μέσω των κρυφών στρωμάτων.

Ο αλγόριθμος εκτελείται σε δύο στάδια. Το πρώτο στάδιο είναι η διάδοση προς τα εμπρός (forward pass), όπου όλα τα δεδομένα εισόδου εισάγονται στο δίκτυο και υπολογίζονται οι έξοδοι. Κατά τη διάρκεια της διαδικασίας αυτής όλες οι ενδιάμεσες τιμές αποθηκεύονται προσωρινά. Το δεύτερο στάδιο περιλαμβάνει την διάδοση προς τα πίσω (backward pass), όπου υπολογίζονται οι παράγωγοι του σφάλματος ως προς τα βάρη του δικτύου. Αυτές οι παράγωγοι χρησιμοπιούνται για να ενημερωθούν τα βάρη μέσω κάποιου αλγορίθμου, όπως ο αλγόριθμος gradient descent. Μετά από αρκετές εμπρόσθιες και οπίσθιες διαδόσεις το δίκτυο μαθαίνει προσαρμόζοντας τα βάρη διαρκώς με σκοπό να ελαχιστοποιηθεί το σφάλμα.

Ο αλγόριθμος gradient descent, ως μέθοδος, εκμεταλλεύεται την παράγωγο της συνάρτησης κόστους με σκοπό να προσαρμόσει τα βάρη με τέτοιο τρόπο έτσι ώστε να φτάσει το κόστος στο ελάχιστο. Αρχικά, προϋποθέτει η συνάρτηση κόστους να είναι παραγωγίσιμη και κυρτή. Ως ξεκίνημα, θέτονται τυχαία βάρη και υπολογίζεται η παράγωγος σε εκείνο το σημείο. Έπειτα, γίνεται ένα βήμα προς την αντίθετη φορά της παραγώγου, δηλαδή προς την φορά που η συνάρτηση κόστους ελαχιστοποιείται. Έτσι, ενημερώνονται τα βάρη με έναν ρυθμό μάθησης (learning rate).

Ο αλγόριθμος gradient descent υπάρχει σε διάφορες παραλλαγές όπως είναι ο στοχαστικός gradient descent, ο οποίος λαμβάνει υπόψιν μόνο ένα τυχαίο δείγμα κάθε φορά, και βρίσκει την παράγωγο για αυτό. Μια άλλη παραλλαγή είναι ο Batch



Σχήμα 2.19: Παράδειγμα πορείας του gradient descent σε σχέση με διάφορα learning rates

Gradient Descent, ο οποίος λαμβάνει υπόψιν όλα τα δεδομένα σε κάθε βήμα και παίρνει τον μέσο όρο των παραγώγων για να ενημερώσει τα βάρη. Τέλος, υπάρχει και ο Mini Batch Gradient Descent, κατά τον οποίο λαμβάνεται υπόψιν ένας προκαθορισμένος αριθμός δειγμάτων, δηλαδή ένα μέρος των δεδομένων.

2.4.5 Συναρτήσεις κόστους (Loss Functions)

Όπως έγινε κατανοητό και από το προηγούμενο υποκεφάλαιο, η συνάρτηση κόστους παίζει σημαντικό ρόλο στο πώς θα εκπαιδευθεί το ΤΝΔ. Επί της ουσίας, μια συνάρτηση κόστους υπολογίζει την διαφορά μεταξύ της προβλεπόμενης και της πραγματικής τιμής. Παρακάτω αναγράφονται οι πιο διαδεδομένες και ευρέως χρησιμοποιούμενες συναρτήσεις κόστους.

- **Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error):** Χρησιμοποιείται σε προβλήματα παλινδρόμησης και υπολογίζεται ως ο μέσος όρος των τετραγώνων των διαφορών μεταξύ πραγματικής και προβλεπόμενης τιμής. Δίνεται από τον τύπο:

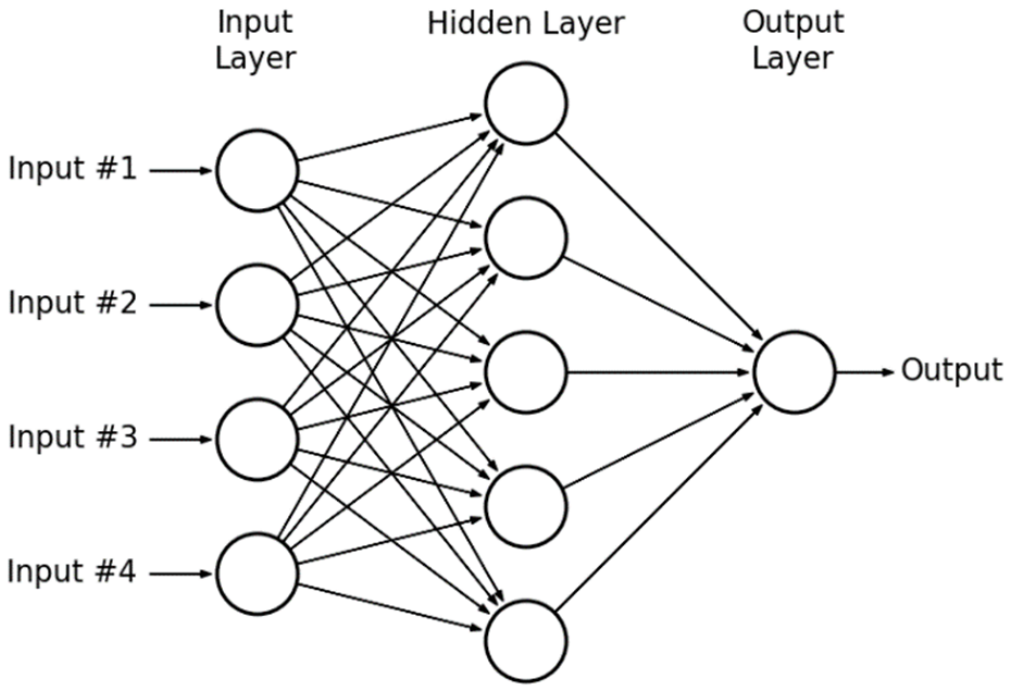
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

  όπου n ο αριθμός των δεδομένων, y_i η πραγματική τιμή και $\hat{y}_i$ η προβλεπόμενη τιμή.

- **Μέσο Απόλυτο Σφάλμα (Mean Absolute Error):** Χρησιμοποιείται επίσης σε προβλήματα παλινδρόμησης και υπολογίζεται ως ο μέσος όρος των απόλυτων διαφορών μεταξύ πραγματικής και προβλεπόμενης τιμής.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- **Huber loss:** Χρησιμοποιείται επίσης για προβλήματα παλινδρόμησης και είναι λιγότερο ευαίσθητη σε ακραίες τιμές.



Σχήμα 2.20: Multi-Layer Perceptron

$$Loss = \begin{cases} \frac{1}{2}(y_i - \hat{y}_i)^2 & \text{αν } |y_i - \hat{y}_i| \leq \delta \\ \delta|y_i - \hat{y}_i| - \frac{1}{2}\delta^2 & \text{αλλιώς} \end{cases}$$

- **Cross Entropy Loss:** Είναι επίσης γνωστή και ως Log Loss και χρησιμοποιείται κατά κόρον σε προβλήματα (δυαδικής) ταξινόμησης. Για να υπολογιστεί, θα πρέπει η έξοδος του ΤΝΔ να είναι κάποια πιθανότητα πρόβλεψης με τιμή από 0 έως και 1.

$$Loss = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right]$$

  Όσο μικρότερη cross-entropy loss έχει ένα μοντέλο, τόσο καλύτερο είναι, αφού η πιθανότητα της πρόβλεψης πλησιάζει στην πραγματική τιμή.

- **Categorical Cross Entropy Loss:** Χρησιμοποιείται για προβλήματα ταξινόμησης πολλαπλών κλάσεων, όπου η έξοδος του ΤΝΔ είναι πολλαπλές κλάσεις.

$$CCE = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{C} y_{ij}\log(\hat{y}_{ij})$$

όπου C είναι ο αριθμός των κλάσεων.

2.4.6 Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNN)

Τα συνελικτικά τεχνητά νευρωνικά δίκτυα είναι ένας τύπος ΤΝΔ που χρησιμοποείται κατά κόρον για μάθηση με τη χρήση εικόνων. Όπως υποδεικνύει και το όνομα τους, χρησιμοποιούν συνέλιξη αντί για πολλαπλασιασμό σε τουλάχιστον ένα κρυφό στρώμα τους. Η συνέλιξη είναι μια πράξη η οποία συνδυάζει μια μικρή περιοχή των δεδομένων με ένα βάρος (ή πυρήνα - kernel) και παράγει έναν άλλο πίνακα. Οι συνελίξεις εφαρμόζονται στα δεδομένα με επαναληπτικό τρόπο και καταφέρνουν να εξάγουν διάφορα χαρακτηριστικά από την εικόνα. Τα CNN είναι ειδικά σχεδιασμένα για να δέχονται δεδομένα με εικονοστοιχεία και βρίσκουν εφαρμογές στην ταξινόμηση εικόνων, στην ανίχνευση προσώπων, καθώς και στην επεξεργασία βίντεο.

Το κύριο δομικό στοιχείο ενός συνελικτικού τεχνητού νευρωνικού δικτύου είναι το στρώμα συνέλιξης (convolution layer). Αυτό το στρώμα υπολογίζει το εσωτερικό γινόμενο δύο πινάκων, δηλαδή μεταξύ του πίνακα παραμέτρων (φίλτρο ή πυρήνας - kernel) οι οποίες ρυθμίζονται κατά την εκπαίδευση, και του πίνακα που υποδεικνύει ένα μέρος της εικόνας, δηλαδή τις τιμές των εικονοστοιχείων της.

Κατά την εμπρόσθια διάδοση, το φίλτρο κινείται κατά μήκος και ύψος της εικόνας και συνελίσσεται, παράγοντας έτσι τον λεγόμενο "χάρτη χαρακτηριστικών" (feature map). Το διάστημα κατά το οποίο κινείται το φίλτρο ονομάζεται stride. Κάθε φορά, το φίλτρο πολλαπλασιάζει τις τιμές των pixel της εικόνας που καλύπτει με τις αντίστοιχες τιμές των βαρών του. Οι πολλαπλασιασμένες τιμές προστίθενται και ο αθροιστής παράγει την τελική τιμή εξόδου.

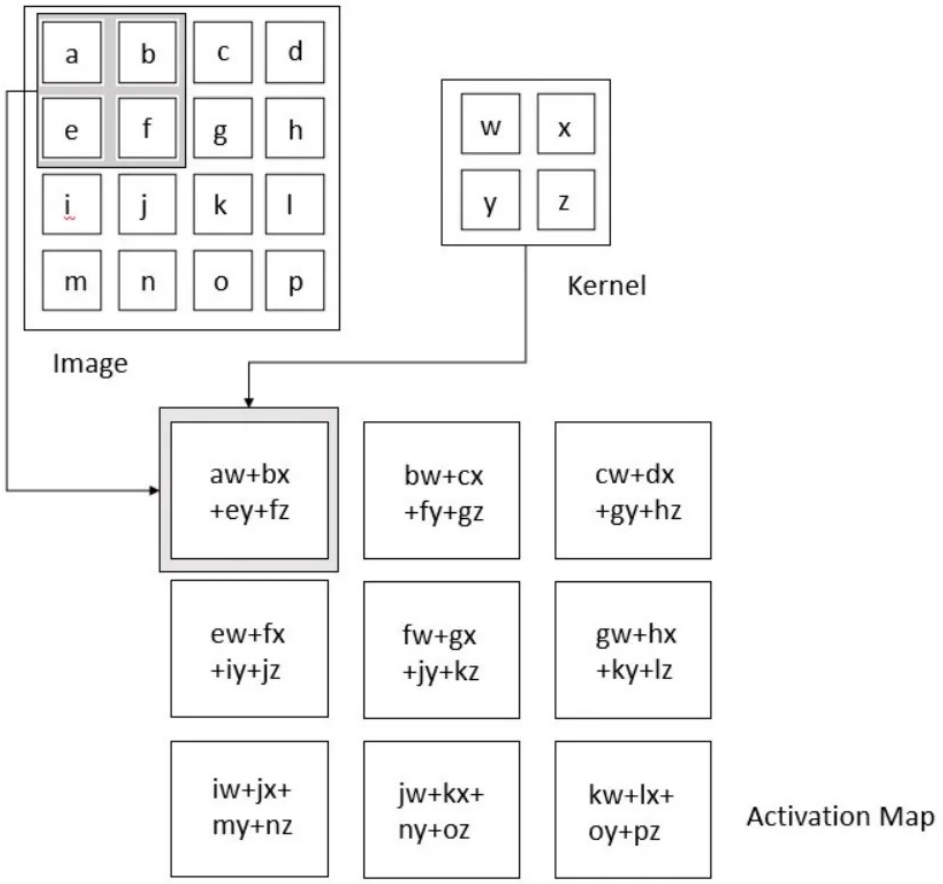
Η συνέλιξη της εικόνας δίνεται από τον παρακάτω τύπο:

$$G[m,n] = (f * h)[m,n] = \sum_j \sum_k h[j,k] f[m-j, n-k]$$

όπου η εικόνα εισόδου συμβολίζεται με f και ο πυρήνας με h, ενώ m και n αφορούν τις γραμμές και τις στήλες που αντιστοιχούν στην περιοχή της εικόνας.

Μετά τη συνέλιξη, εφαρμόζεται μια συνάρτηση ενεργοποίησης (όπως η σιγμοειδής ή η ReLU) στον χάρτη χαρακτηριστικών για να εισαγάγει μη γραμμικότητα στο δίκτυο. Στην συνέχεια η διαδικασία της συνέλιξης επαναλαμβάνεται για όλα τα φίλτρα σε ένα στρώμα, παράγοντας έναν νέο χάρτη χαρακτηριστικών για κάθε φίλτρο. Αυτός ο νέος χάρτης χαρακτηριστικών είναι η είσοδος στο επόμενο στρώμα του CNN, όπου η διαδικασία επαναλαμβάνεται. Μέσω αυτής της διαδικασίας, το CNN είναι σε θέση να εξάγει χαρακτηριστικά από την εικόνα.

Ένα κύριο πλεονέκτημά τους είναι πως μπορούν να αναγνωρίσουν τοπικά χα-

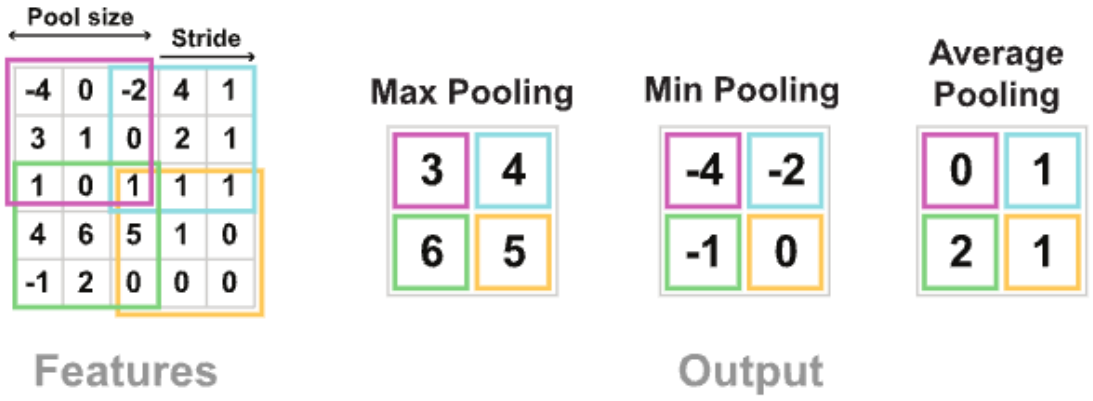


Σχήμα 2.21: Γραφική αναπαράσταση της διαδικασίας συνέλιξης

ρακτηριστικά σε μια εικόνα, χωρίς να επηρεάζονται από την θέση τους. Είναι, δηλαδή, ανθεκτικά στην μετατόπιση, την αλλοίωση και την αλλαγή κλίμακας των δεδομένων εισόδου. Επίσης, τα CNN περιέχουν συνήθως στρώματα που εκτελούν υπο-δειγματοληψία (pooling layers), τα οποία μειώνουν το μέγεθος των χαρακτηριστικών, χωρίς να χάνεται σημαντική πληροφορία.

Υπάρχουν διάφορες τεχνικές για να εφαρμοστεί υπο-δειγματοληψία, όπως εξηγείται και στην εικόνα 2.22. Για παράδειγμα, ξεκινώντας από έναν πίνακα 5×5 μπορούμε να πάμε σε έναν πίνακα 2×2 χωρίζοντας αρχικά τον πίνακα σε 4 υπο-πίνακες, ο καθένας από τους οποίους σημειώνεται με ένα μοναδικό χρώμα. Ο διαχωρισμός αυτός γίνεται αρχικά βάσει των παραμέτρων pool size, δηλαδή πόσο μεγάλος θα είναι ο υπο-πίνακας, και stride, δηλαδή με ποιο βήμα θα κινείται μέσα στην εικόνα. Εν συνεχεία, για καθέναν από τους υπο-πίνακες βρίσκεται είτε το μέγιστο ή ελάχιστο στοιχείο, είτε ο μέσος όρος των στοιχείων. Έχουμε, δηλαδή max pooling, min pooling και average pooling αντίστοιχα.

Το μέγεθος του πίνακα που θα προκύψει μετά από το pooling δίνεται από τον παρακάτω τύπο:



Σχήμα 2.22: Είδη στρωμάτων Pooling

$$W_{out} = \frac{W - F}{S} + 1$$

όπου W το μέγεθος του πίνακα, F το μέγεθος του φίλτρου pooling και S το stride, δηλαδή το βήμα που κινείται πάνω στον πίνακα. Ακολουθώντας το παράδειγμα της εικόνας 2.22, ένας πίνακας 5×5 θα γίνει 2×2 με stride 2 και φίλτρο 3×3 γιατί

$$W_{out} = \frac{5 - 3}{2} + 1 = 2$$

2.4.7 Επαναλαμβανόμενα Νευρωνικά Δίκτυα (Recurrent Neural Networks - RNN)

Τα επαναλαμβανόμενα τεχνητά νευρωνικά δίκτυα είναι ένας τύπος ΤΝΔ που χρησιμοποιούνται κυρίως για επεξεργασία και μάθηση δεδομένων σε μορφή ακολουθίας, όπως οι χρονολογικές σειρές (time series). Σε αντίθεση με τα κλασικά νευρωνικά δίκτυα τα οποία λαμβάνουν υπόψιν την είσοδο και παράγουν έξοδο για τη δεδομένη χρονική στιγμή, τα επαναλαμβανόμενα ΤΝΔ επιτρέπουν την επεξεργασία της εισόδου με βάση την πληροφορία που έχει αποθηκευθεί από προηγούμενες χρονικές στιγμές. Ουσιαστικά, επιτρέπουν στην έξοδο ενός νευρώνα να επηρεάσει την είσοδο που ακολουθεί στους ίδιους νευρώνες.

Τα RNN βρίσκουν εφαρμογή σε προβλήματα που σχετίζονται με δεδομένα σε ακολουθιακή μορφή, όπως η αναγνώριση και πρόβλεψη της τιμής μίας μετοχής, η αναγνώριση φωνής και η αυτόματη μετάφραση.

2.4.8 Δίκτυα μακράς βραχυπρόθεσμης μνήμης (Long Short-Term Memory networks - LSTM)

Τα δίκτυα μακράς βραχυπρόθεσμης μνήμης (Long Short-Term Memory (LSTM)) είναι μια παραλλαγή των επαναλαμβανόμενων νευρωνικών δικτύων η οποία προτάθηκε το 1997 από τους Sepp Hochreiter και Jürgen Schmidhuber [20]. Τα LSTM

είναι εξίσου ιδανικά για την μάθηση δεδομένων σε μορφή ακολουθίας και είναι ικανά να μάθουν μακροπρόθεσμες εξαρτήσεις σε αυτά. Σε αντίθεση με τα παραδοσιακά RNN, τα LSTM έχουν μια πρόσθετη κυψέλη μνήμης (memory cell) η οποία τους επιτρέπει να διατηρούν πληροφορίες για περισσότερα χρονικά βήματα.

Τα LSTM αποτελούνται από το κελί μνήμης (memory cell), μία πύλη εισόδου (input gate), μία πύλη εξόδου (output gate) και μία πύλη λήθης (forget gate). Το κελί μνήμης "θυμάται" δεδομένα για ένα απροσδιόριστο χρονικό διάστημα και οι τρεις προαναφερθείσες πύλες ελέγχουν την πληροφορία από και προς αυτό. Οι πύλες λήθης αποφασίζουν ποιά πληροφορία θα ξεχαστεί από τη μνήμη, ενώ η πύλη εισόδου αποφασίζει αντίστοιχα τι θα αποθηκευτεί. Τέλος, η πύλη εξόδου αποφασίζει ποιες πληροφορίες θα βγουν από το κελί μνήμης.

Τα LSTM αποτελούνται από:

- **Πύλη λήθης (forget gate):** Η πύλη λήθης αποφασίζει ποια πληροφορία θα διαγραφεί από τη μνήμη. Αυτό είναι χρήσιμο για να διαγραφούν από την μνήμη οι πληροφορίες που δεν είναι πλέον χρήσιμες ή μπορεί να είναι άσχετες με την τρέχουσα ακολουθία δεδομένων.

$$f_t = \sigma(W_i x_t + U_f h_{t-1} + b_f)$$

όπου h_{t-1} η έξοδος της προηγούμενης χρονικής στιγμής.

- **Πύλη εισόδου (input gate):** Οι πύλες εισόδου αποφασίζουν ποιες πληροφορίες θα αποθηκευτούν στην μνήμη με βάση την προηγούμενη τιμή εξόδου και την νέα τιμή εισόδου.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

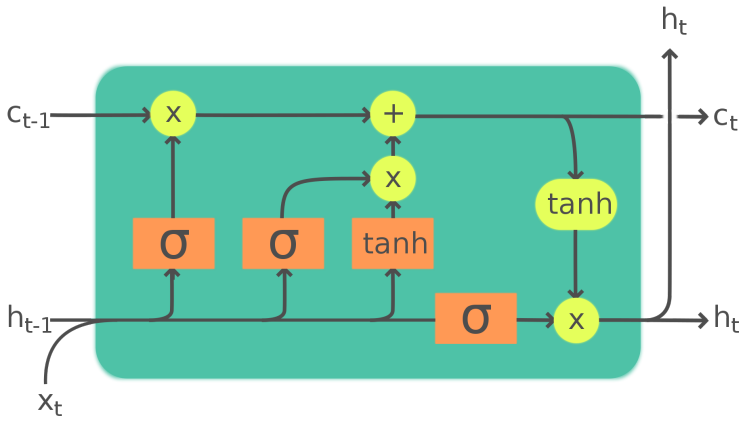
$$\tilde{c_x} = tanh(W_c X_t + U_c h_{t-1} + b_c)$$

- **Πύλη εξόδου (output gate):** Οι πύλες εξόδου καθορίζουν ποιες πληροφορίες θα εξέλθουν από το LSTM και προωθούνται προς την επόμενη μονάδα.

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$

$$h_t = o_t \cdot tanhc_t(\text{hidden state})$$

- **Κελί μνήμης (memory cell):** Το κελί μνήμης χρησιμοποιείται για να αποθηκεύει και να επιλέγει πληροφορίες που είναι σημαντικές για μελλοντική χρήση.



Σχήμα 2.23: Απεικόνιση δικτύου LSTM

$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t$$

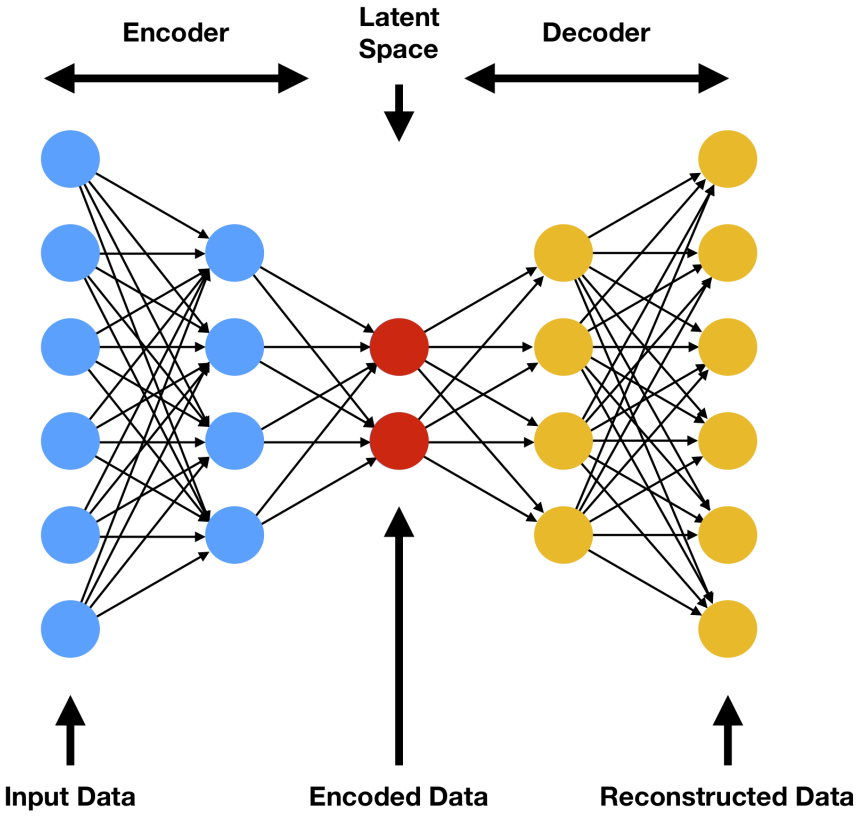
2.4.9 Autoencoders

Ένας autoencoder [1] αποτελεί ένα ΤΝΔ το οποίο χρησιμοποιείται για μη επιβλεπόμενη μάθηση. Πιο συγκεκριμένα, χρησιμοποιείται για την εκμάθηση δεδομένων και την μείωση της διάστασης των εισερχομένων δεδομένων σε έναν χαμηλότερο χώρο. Οι autoencoders αποτελούνται από δύο κύρια μέρη: τον κωδικοποιητή (encoder) και τον αποκωδικοποιητή (decoder). Ο κωδικοποιητής αναλαμβάνει τη μείωση της διάστασης των εισερχομένων δεδομένων σε έναν χαμηλότερο χώρο, ενώ ο αποκωδικοποιητής αναλαμβάνει την αποκωδικοποίηση αυτής της αναπαράστασης πίσω στον αρχικό χώρο.

Στα απλά νευρωνικά δίκτυα, η διαδικασία της εκπαίδευσης γίνεται έχοντας ως στόχο τις ετικέτες των δεδομένων (π.χ. σε μια δυαδική ταξινόμηση e-mail οι ετικέτες είναι spam ή not spam. Αντιθέτως, οι autoencoders δεν δέχονται ετικέτες ως στόχο, αλλά αναδημιουργήσουν τα ήδη υπάρχοντα δεδομένα με μια πιο περιορισμένη και συμπιεσμένη αναπαράσταση, αντλώντας έτσι τις κύριες χαρακτηριστικές πληροφορίες από την είσοδο.

Οι autoencoders χρησιμοποιούνται σε πολλούς τομείς, όπως η μείωση της διάστασης των δεδομένων και η αναγνώριση προτύπων. Επίσης, αποτελούν σημαντικό εργαλείο για την ανάκτηση πληροφορίας από ακουστικά ή οπτικά σήματα.

[1] Στα ελληνικά συναντάται σπάνια και ως "αυτοκωδικοποιητής", αλλά η αγγλική ορολογία είναι η επικρατέστερη



Σχήμα 2.24: Η δομή ενός autoencoder

2.5 Transfer Learning

Το transfer learning αποτελεί μια ισχυρή τεχνική στον χώρο της μηχανικής μάθησης και της τεχνητής νοημοσύνης. Η ουσία του βασίζεται στην ιδέα ότι τα νευρωνικά δίκτυα που έχουν εκπαιδευτεί για ένα πρόβλημα, μπορούν να μεταφερθούν και να χρησιμοποιηθούν για να λύσουν άλλα προβλήματα. Αντί να εκπαιδεύει ένα νέο μοντέλο από την αρχή, το transfer learning χρησιμοποιεί την προηγούμενη εκπαίδευση ως αφετηρία και προσπαθεί να προσαρμόσει τη μάθηση στο νέο πρόβλημα. Η μέθοδος αυτή αξιοποιείται σε διάφορους κλάδους, όπως η ταξινόμηση κειμένου, η ρομποτική, ή ακόμα και η υγεία.

Ένα από τα βασικά πλεονεκτήματα του transfer learning είναι η εξοικονόμηση χρόνου και πόρων. Αντί να ξεκινήσουμε από το μηδέν και να εκπαιδεύσουμε ένα μοντέλο από την αρχή, μπορούμε να ξεκινήσουμε με ένα προεκπαιδευμένο μοντέλο που έχει ήδη μάθει πολλά χαρακτηριστικά από δεδομένα παρόμοια με αυτά που μας αφορούν. Από εκεί, μπορούμε να προσαρμόσουμε το μοντέλο για να λύσει το συγκεκριμένο πρόβλημα.

Το transfer learning εφαρμόζεται πολύ συχνά στον τομέα της υπολογιστικής όρασης. Υπάρχουν αρκετά προεκπαιδευμένα μοντέλα τα οποία χρησιμοποιούνται ευρέως για αναγνώριση εικόνων, ταξινόμηση αντικειμένων ή ακόμα και τμηματοποίηση εικόνων. Ένα από αυτά είναι το VGG [21], το οποίο πέτυχε 92.7 % ορθότητα στο σετ εικόνων ImageNet το οποίο περιέχει 14 εκατομμύρια εικόνες 1000 κλάσεων. Ένα άλλο γνωστό μοντέλο είναι το AlexNet [22], το οποίο είναι ένα CNN 62.3 εκατομμυρίων παραμέτρων. Τέτοια προεκπαιδευμένα τεχνητά νευρωνικά δίκτυα, και ειδικότερα CNN, έχουν εφαρμοσθεί σε ποικίλα πεδία όπως η ανίχνευση κακόβουλων bots, ή αλλιώς spambots, στο Twitter [23], η ανίχνευση άνοιας από την ομιλία [24], και η ανίχνευση της επιληψίας [25].

2.6 Επιλογή Χαρακτηριστικών

Η επιλογή χαρακτηριστικών (feature selection) αναφέρεται στην διαδικασία κατά την οποία επιλέγεται ένα υποσύνολο χαρακτηριστικών από το σετ δεδομένων. Η επιλογή αυτή γίνεται βάσει της πληροφορίας της οποίας παρέχουν στο πρόβλημα προς επίλυση. Αν επιλεχθεί το σωστό σετ χαρακτηριστικών, τότε ο χρόνος μάθησης μειώνεται, ενώ η απόδοση αυξάνεται. Ταυτόχρονα, μπορεί να αποφευχθεί το overfitting και να βελτιωθεί η ερμηνευσιμότητα του μοντέλου. Ένα μοντέλο ορίζεται ως ερμηνεύσιμο όταν οι προβλέψεις του μπορούν να ερμηνευθούν από τους ανθρώπους και μπορεί να κατανοηθεί η αιτία της κάθε πρόβλεψης.

Οι μέθοδοι επιλογής χαρακτηριστικών μπορούν να χωριστούν σε τρεις κατηγορίες: τις μεθόδους φιλτραρίσματος (filter methods), τις ενσωματωμένες μεθόδους (embedded methods) και τις μεθόδους περιτυλίγματος (wrapper methods).

2.6.1 Μέθοδοι Φιλτραρίσματος (Filter methods)

Οι μέθοδοι φιλτραρίσματος αξιολογούν την σχέση των χαρακτηριστικών χωρίς να λαμβάνεται υπόψιν το μοντέλο μηχανικής μάθησης. Εν αντιθέσει, χρησιμοποιείται η τιμή διαφόρων στατιστικών τεστ για να αποφασιστεί η σχέση τους με την τιμή εξόδου.

Παρακάτω αναφέρονται κάποια από αυτά τα στατιστικά τεστ τα οποία χρησιμοποιούνται σε αυτή την περίπτωση:

- **Συσχέτιση Pearson:**

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

- **Συσχέτιση Spearman:**

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

- **Αμοιβαία Πληροφορία (Mutual Information):** Είναι μια μη αρνητική τιμή η οποία μετράει την εξάρτηση δύο μεταβλητών. Η αμοιβαία πληροφορία δύο διακριτών μεταβλητών X και Y δίνεται από τον παρακάτω τύπο.

$$I(X;Y) = \sum_{x \in X}\sum_{y \in Y} P(x,y)\log\left(\frac{P(x,y)}{P(x)P(y)}\right)$$

- **Chi-squared test (χ^2):** Χρησιμοποιείται για να αξιολογήσει αν υπάρχει στατιστικά σημαντική συσχέτιση μεταξύ δύο ποιοτικών μεταβλητών. Είναι επίσης γνωστό ως τεστ ανεξαρτησίας χ^2.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

  όπου O είναι οι συχνότητες που παρατηρούνται και E οι αναμενόμενες συχνότητες των μεταβλητών.

- **Relief [26]:** Πρόκειται για μια μέθοδο φιλτραρίσματος η οποία αρχικά επιλέγει ένα τυχαίο δείγμα από τα δεδομένα και έπειτα υπολογίζει τις αποστάσεις μεταξύ αυτού του δείγματος και όλων των άλλων δειγμάτων στο σύνολο δεδομένων. Οι αποστάσεις υπολογίζονται τόσο για περιπτώσεις της ίδιας κλάσης (θετικές περιπτώσεις) όσο και για περιπτώσεις διαφορετικών κλάσεων (αρνητικές περιπτώσεις). Για κάθε χαρακτηριστικό του δείγματος, ο αλγόριθμος ενημερώνει το βάρος με βάση τις διαφορές μεταξύ των τιμών των χαρακτηριστικών του και των πλησιέστερων θετικών (near hits) και αρνητικών δειγμάτων (near misses)

για αυτό το χαρακτηριστικό. Ο στόχος είναι να αυξηθεί το βάρος των χαρακτηριστικών που συμβάλλουν στη σωστή ταξινόμηση (έχουν παρόμοιες τιμές με κοντινά θετικά δείγματα) και να μειωθεί το βάρος για τα χαρακτηριστικά που προκαλούν εσφαλμένη ταξινόμηση (έχουν διαφορετικές τιμές από κοντινά αρνητικά δείγματα).

Η διαδικασία αυτή επαναλαμβάνεται είτε για έναν συγκεκριμένο αριθμό επαναλήψεων, είτε μέχρι να έχουν ληφθεί υπόψιν όλα τα δείγματα. Μετά την ολοκλήρωση των επαναλήψεων, ο αλγόριθμος ταξινομεί τα χαρακτηριστικά με βάση τα συνολικά βάρη τους. Τα χαρακτηριστικά με μεγαλύτερα βάρη θεωρούνται πιο σχετικά για την ταξινόμηση, ενώ αυτά με χαμηλότερα βάρη θεωρούνται λιγότερο σημαντικά. Έπειτα, με βάση την κατάταξη των χαρακτηριστικών, ο αλγόριθμος μπορεί στη συνέχεια να επιλέξει ένα υποσύνολο χαρακτηριστικών που θα χρησιμοποιηθεί για την κατασκευή του μοντέλου ταξινόμησης. Ο χρήστης μπορεί να αποφασίσει τον αριθμό των χαρακτηριστικών που θα διατηρηθούν, όπως για παράδειγμα τα δέκα πρώτα, τα πέντε πρώτα κ.ο.κ.

Ο αλγόριθμος Relief είναι επωφελής, καθώς μπορεί να χειριστεί τόσο συνεχή όσο και χαρακτηριστικά με κατηγορίες και λαμβάνει υπόψη τις αλληλεπιδράσεις μεταξύ των χαρακτηριστικών. Ωστόσο, απαιτεί πολλές επαναλήψεις, καθιστώντας τον υπολογιστικά ακριβό για μεγάλα σύνολα δεδομένων.

- **ReliefF [27]:** Πρόκειται για μια βελτιωμένη μορφή του Relief [26], η οποία υπολογίζει τις αποστάσεις από τα κοντινότερα δείγματα χρησιμοποιώντας την απόσταση Manhattan και την Ευκλείδια νόρμα, αλλά και χρησιμοποιεί την απόλυτη διαφορά των near hits και near misses ως οδηγό για την αλλαγή των βαρών. Ο αλγόριθμος ReliefF εφαρμόζεται επίσης σε λειψά δεδομένα.

  Ο αλγόριθμος Relief συναντάται σε πολλές παραλλαγές, όπως ο RRELIEFF [28], ο Relieved-F [29], ο TuRF (Tuned ReliefF) [30] και ο SURF (Spatially uniform relieff) [31].

Οι μέθοδοι φιλτραρίσματος είναι πιο γρήγορες από άποψη πολυπλοκότητας χρόνου και είναι λιγότερο επιρρεπείς στο overfitting.

2.6.2 Ενσωματωμένες Μέθοδοι (Embedded methods):

Οι ενσωματωμένες μέθοδοι εκμεταλλεύονται την διαδικασία εκπαίδευσης του μοντέλου για την επιλογή των χαρακτηριστικών. Αντί να εξετάζουν απλώς την σχέση μεταξύ του στόχου και των χαρακτηριστικών μεμονωμένα, οι μέθοδοι αυτές χρησιμοποιούν πληροφορίες από την ίδια την διαδικασία της εκπαίδευσης για να αξιολογήσουν την σημασία ενός χαρακτηριστικού.

Ένα από τα μοντέλα τα οποία έχουν από μόνα τους την δυνατότητα να επιλέγουν χαρακτηριστικά είναι τα δέντρα αποφάσεων και τα άλλα μοντέλα που βασίζονται σε

αυτά, όπως τα τυχαία δάση (Random Forests). Κατά την εκπαίδευση ενός δέντρου αποφάσεων αξιολογείται η σημασία παρατηρώντας πόσο ένα χαρακτηριστικό μειώνει την λανθασμένη ταξινόμηση, δηλαδή πόσο πληροφοριακό κέρδος δίνει. Αν το κέρδος είναι μεγάλο, τότε το χαρακτηριστικό θεωρείται σημαντικό.

Άλλες μέθοδοι περιλαμβάνουν μεθόδους κανονικοποίησης όπως η Lasso (L1 Regularization) και η Ridge (L2 Regularization). Η κανονικοποίηση Lasso (Least Absolute Shrinkage and Selection Operator) ή L1 Regularization, προσθέτει έναν όρο ποινής στην συνάρτηση κόστους. Ο όρος αυτός ονομάζεται απόλυτη τιμή μεγέθους (absolute value of magnitude) και εισάγει έναν όρο περιορισμού στο άθροισμα των απόλυτων τιμών των συντελεστών. Ως αποτέλεσμα, ορισμένοι συντελεστές συμπιέζονται προς το μηδέν και έτσι προκύπτει ένα αραιό μοντέλο με μικρότερο αριθμό σημαντικών χαρακτηριστικών. Ο όρος περιορισμού ή όρος ποινής είναι ίσος με

$$\lambda \sum_{i=1}^{n} |\beta_i|$$

όπου λ είναι η υπερπαράμετρος που ελέγχει το επίπεδο της ποινής και το βαθμό συρρίκνωσης των συντελεστών, n είναι ο αριθμός των συντελεστών (βαρών) του μοντέλου και β_i i-οστός συντελεστής (βάρος) του μοντέλου.

Η κανονικοποίηση Lasso βοηθά στην αυτόματη επιλογή των πιο σημαντικών χαρακτηριστικών και εξαλείφει τα ανεπιθύμητα χαρακτηριστικά, μειώνοντας την πολυπλοκότητα του μοντέλου και αποτρέποντας την υπερεκτίμηση. Έτσι, επιτυγχάνεται μια καλύτερη γενίκευση στα νέα δεδομένα.

Η κανονικοποίηση Ridge (L2 Regularization), από την άλλη, προσθέτει έναν άλλο όρο ποινής ο οποίος είναι ίσος με

$$\lambda \sum_{i=1}^{n} \beta_i^2$$

Με την εισαγωγή του όρου περιορισμού, οι μεγάλες τιμές των συντελεστών περιορίζονται, ενώ οι μικρές τιμές διατηρούνται. Αυτό έχει ως αποτέλεσμα τη μείωση της ευαισθησίας του μοντέλου στον θόρυβο και τις ασυνήθιστες τιμές (outliers) των δεδομένων. Εφαρμόζεται συνήθως σε γραμμικά μοντέλα, όπως η γραμμική παλινδρόμηση, ενώ επεκτείνεται επίσης και σε άλλες μεθόδους, όπως τα νευρωνικά δίκτυα.

Και οι δύο μέθοδοι κανονικοποίησης χρησιμοποιούνται επίσης για να αποφευχθεί το overfitting. Η βασική διαφορά τους είναι ότι η κανονικοποίηση Lasso μειώνει τους συντελεστές των λιγότερο σημαντικών χαρακτηριστικών στο μηδέν. Εν αντιθέσει, κατά την κανονικοποίηση Ridge τα βάρη τείνουν προς το μηδέν, χωρίς όμως να φτάσουν υποχρεωτικά σε αυτό. Έτσι, η επίδρασή τους στο μοντέλο μειώνεται.

Οι ενσωματωμένες μέθοδοι είναι πιο αργές σε σχέση με τις μεθόδους φιλτραρίσματος. Ωστόσο, είναι πιο γρήγορες από άποψη πολυπλοκότητας χρόνου από τις μεθόδους περιτυλίγματος, οι οποίες παρουσιάζονται στην αμέσως επόμενη υποενότητα. Συνήθως χρησιμοποιούνται για να μειώσουν το overfitting, προσθέτοντας όρους ποινής, όπως ακριβώς κάνουν οι μέθοδοι κανονικοποίησης Lasso και Ridge.

2.6.3 Μέθοδοι Περιτυλίγματος (Wrapper methods)

Οι μέθοδοι περιτυλίγματος ή wrapper methods [32] αφορούν μεθόδους οι οποίες επιλέγουν τα χαρακτηριστικά βάσει της απόδοσης του μοντέλου μηχανικής μάθησης. Εκτελούν επαναληπτικά την διαδικασία εκπαίδευσης και αξιολογούν την επίδοση του μοντέλου για διάφορους συνδυασμούς χαρακτηριστικών. Οι μέθοδοι περιτυλίγματος περιλαμβάνουν τις μεθόδους εμπρόσθιας επιλογής χαρακτηριστικών (step forward feature selection), τις μεθόδους οπίσθιας επιλογής χαρακτηριστικών (step backward feature selection), τις μεθόδους εξαντλητικής επιλογής χαρακτηριστικών (exhaustive feature selection), καθώς και τις μεθόδους αναδρομικής εξάλειψης χαρακτηριστικών recursive feature elimination.

- **Εμπρόσθια επιλογή χαρακτηριστικών (step forward feature selection):** Κατά την εμπρόσθια επιλογή χαρακτηριστικών, ο αλγόριθμος ξεκινάει με ένα "άδειο" μοντέλο, προσθέτοντας ένα μεμονωμένο χαρακτηριστικό κάθε φορά και επιλέγοντας την χαμηλότερη τιμή p (p-value). Έπειτα, εκπαιδεύεται ένα μοντέλο προσθέτοντας άλλο ένα χαρακτηριστικό κάθε φορά, το οποίο επιλέγεται με τον ίδιο τρόπο. Αυτή η διαδικασία επαναλαμβάνεται έως ότου προκύψει ένα σετ χαρακτηριστικών τα οποία έχουν p-values χαμηλότερα από το προεπιλεγμένο επίπεδο σημαντικότητας.

- **Οπίσθια επιλογή χαρακτηριστικών (step backward feature selection):** Η οπίσθια επιλογή χαρακτηριστικών γίνεται κατά τρόπο αντίθετο από την εμπρόσθια μέθοδο. Σε αυτή την περίπτωση, ο αλγόριθμος ξεκινάει με ένα μοντέλο που περιέχει όλα τα χαρακτηριστικά και κάθε φορά "πετάει" τη μεγαλύτερη τιμή p (p-value). που είναι πάνω από τα όρια του επιπέδου σημαντικότητας.

- **Εξαντλητική επιλογή χαρακτηριστικών (Exhaustive feature selection):** Η εξαντλητική επιλογή χαρακτηριστικών αξιολογεί κάθε πιθανό σετ χαρακτηριστικών λαμβάνοντας υπόψιν όλους τους συνδυασμούς, από το πιο μικρό σετ (δηλαδή μόνο ένα χαρακτηριστικό), μέχρι το πιο μεγάλο (δηλαδή το σετ που περιέχει όλα τα χαρακτηριστικά).

- **Αναδρομική εξάλειψη χαρακτηριστικών (recursive feature elimination):** Η αναδρομική εξάλειψη χαρακτηριστικών ακολουθεί μια συστηματική επαναληπτική διαδικασία για την εύρεση του καλύτερου σετ χαρακτηριστικών. Εφαρμόσθηκε πρώτη φορά από τους Guyon et al. [33] το 2002. Ο αλγόριθμος ξεκινάει χρησιμοποιώντας όλα τα χαρακτηριστικά και εκπαιδεύει ένα μοντέλο (συνήθως κάποιο μοντέλο βασισμένο σε τυχαία δέντρα ή ένα μοντέλο γραμμι-

κής παλινδρόμησης). Στη συνέχεια, ανατίθεται σε κάθε χαρακτηριστικό ένας βαθμός σημαντικότητας ή αριθμός κατάταξης, ο οποίος μπορεί να προέλθει από το ίδιο το μοντέλο, και τα λιγότερο σημαντικά χαρακτηριστικά εξαλείφονται. Ο αριθμός των χαρακτηριστικών που πρέπει να εξαλειφθούν σε κάθε επανάληψη συνήθως ορίζεται από τον χρήστη.

Το επόμενο βήμα είναι η επανεκπαίδευση του μοντέλου με το σετ χαρκτηριστικών που προέκυψε. Ο στόχος είναι να αξιολογηθεί πώς αλλάζει η απόδοση του μοντέλου μετά την αφαίρεση των λιγότερο σημαντικών χαρακτηριστικών. Η επανάληψη συνεχίζεται, εξαλείφοντας σταδιακά τα χαρακτηριστικά και επανεκπαιδεύοντας το μοντέλο, μέχρι να ικανοποιηθεί ένα κριτήριο διακοπής, το οποίο επίσης ορίζεται από τον χρήστη. Αυτό το κριτήριο θα μπορούσε να είναι ένας προκαθορισμένος αριθμός επαναλήψεων, η επίτευξη ενός επιθυμητού αριθμού χαρακτηριστικών ή ένα συγκεκριμένο όριο απόδοσης (π.χ. ορθότητα για ταξινόμηση ή R^2 για παλινδρόμηση).

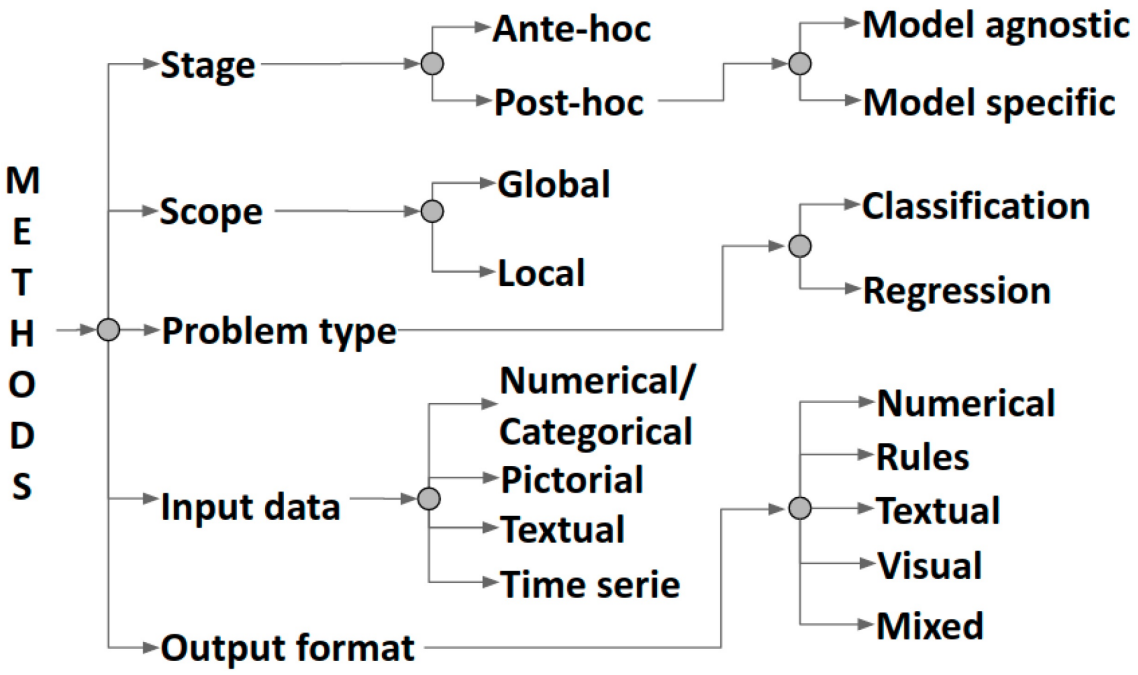
Οι μέθοδοι περιτυλίγματος είναι γενικά περισσότερο επιρρεπείς στο overfitting, καθώς εκπαιδεύουν μοντέλα με διάφορους συνδυασμούς χαρακτηριστικών. Είναι επίσης πιο αργές, ειδικά για δεδομένα με πολλά χαρακτηριστικά.

2.7 Ερμηνευσιμότητα μοντέλων - Explainable AI

Τα μοντέλα μηχανικής μάθησης θα μπορούσαν να χωριστούν σε δύο κατηγορίες: τα "white box" και τα "black box" μοντέλα [3]. Ως "black box" μοντέλα νοούνται τα μοντέλα τα οποία είναι σαν ένα "μαύρο κουτί", δηλαδή δεν είναι "διαφανή" και είναι πολύ δύσκολο να κατανοηθούν και να εξηγηθούν. Πολλές φορές, όταν ένα μοντέλο καταλήγει σε μια πρόβλεψη, υπάρχει η ανάγκη να εξηγηθεί ο τρόπος με τον οποίο παρήχθη αυτή η πρόβλεψη.

Ο τομέας ο οποίος ασχολείται με αυτό το πρόβλημα ονομάζεται Explainable AI (XAI) και ο βασικός του στόχος είναι η ανάπτυξη μοντέλων που μπορούν να εξηγήσουν τις λήψεις αποφάσεων που πραγματοποιούν, δίνοντας στους χρήστες τη δυνατότητα να κατανοήσουν τον λόγο πίσω από τις προβλέψεις τους. Η ανάγκη για το Explainable AI προέκυψε από την αυξημένη χρήση μηχανικής μάθησης σε κρίσιμες εφαρμογές, όπως η ιατρική διάγνωση, τα αυτοκίνητα αυτόνομης οδήγησης και η δικαιοσύνη, όπου η διαφάνεια και η αιτιολογησιμότητα των αποφάσεων είναι κρίσιμης σημασίας. Ειδικότερα σε εφαρμογές της τεχνητής νοημοσύνης που επηρεάζουν άμεσα τους ανθρώπους, όπως οι εφαρμογές στην δικαιοσύνη και το νομικό σύστημα, είναι σημαντικό να υπάρχει και η επεξήγηση της πρόβλεψης [34]. Ο σκοπός είναι να επιβεβαιωθεί πως το μοντέλο δεν μεροληπτεί υπέρ κάποιας κοινωνικής ομάδας, επηρεάζοντας έτσι αρνητικά κάποια άλλη.

Οι Vilone και Longo [3] ιεράρχησαν τις διαθέσιμες μεθόδους Explainable AI με βάση διάφορα κριτήρια, όπως ο τύπος προβλήματος (ταξινόμηση ή παλινδρόμηση), τα δεδομένα εισόδου (εικόνα, κείμενο, αριθμητικά δεδομένα κ.α.), ο τύπος εξόδου, το πλαίσιο εφαρμογής και το στάδιο (πριν ή μετά την εκπαίδευση).



Σχήμα 2.25: Ιεράρχηση των μεθόδων XAI κατά τις Vilone και Longo [3]

Οι μέθοδοι Explainable AI περιλαμβάνουν την ανάλυση των χαρακτηριστικών του μοντέλου, την ερμηνεία των αποφάσεών τους με τη χρήση γραφικών μοντέλων, την ανάπτυξη μοντέλων επεξήγησης που προσεγγίζουν τη λειτουργία του μοντέλου μηχανικής μάθησης, καθώς και τη χρήση διαφορετικών μεθόδων όπως το LIME (Local Interpretable Model-agnostic Explanations) [35] και το SHAP (SHapley Additive exPlanations) [36].

Το LIME [35] καλείται να δώσει απάντηση στην ερώτηση γιατί το μοντέλο προβλέπει μία συγκεκριμένη τιμή και βοηθά στην κατανόηση των αποφάσεων που παίρνει ένα μοντέλο μηχανικής μάθησης, αποκαλύπτοντας τα σημαντικότερα χαρακτηριστικά που επηρεάζουν τις προβλέψεις του μοντέλου. Μέσω του LIME τροποποιείται ένα μεμονωμένο δείγμα δεδομένων το οποίο με την σειρά του τροποποιεί τις τιμές των χαρακτηριστικών και παρατηρεί αντίκτυπο που αντανακλάται στην έξοδο. Στη συνέχεια, υπολογίζεται η σημασία κάθε χαρακτηριστικού, δίνοντας μια εξήγηση για τον τρόπο με τον οποίο το μοντέλο μεταβάλλει τις προβλέψεις βάσει των τιμών των χαρακτηριστικών.

Το LIME έχει χρησιμοποιηθεί ευρέως σε ερευνητικές εργασίες, όπως για παράδειγμα στην επεξήγηση της πρόβλεψης της εξαπάτησης μέσω κειμένου (text-based deception) [37], ή στην ερμηνεία της προβλεψης της άνοιας από απομαγνητοφωνήσεις [38].

Από την άλλη, η μέθοδος SHAP [36] βασίζεται στην ανάλυση της συνεισφοράς κάθε χαρακτηριστικού στην τελική πρόβλεψη, θέτοντας τον Συντελεστή Shapley (Shapley value) ως μέτρο σημαντικότητας κάθε χαρακτηριστικού. Η μέθοδος SHAP βασίζεται στην θεωρία παιγνίων (game theory). Ωστόσο, θα πρέπει να χρησιμοποιείται με

σύνεση καθώς δείχνει την σημαντικότητα κάθε χαρακτηριστικού αλλά δεν αξιολογεί την ποιότητα της πρόβλεψης.

Το SHAP προσφέρει διάφορες λειτουργίες εκτός από την ανάλυση της συνεισφοράς, όπως είναι διάφορες γραφικές απεικονίσεις των τιμών. Επίσης, εκτός από μοντέλα που δέχονται ως είσοδο δεδομένα σε μορφή πίνακα, χρησιμοποιείται και σε διάφορα μοντέλα, όπως μοντέλα υπολογιστικής όρασης τα οποία χρησιμοποιούν εικόνες.

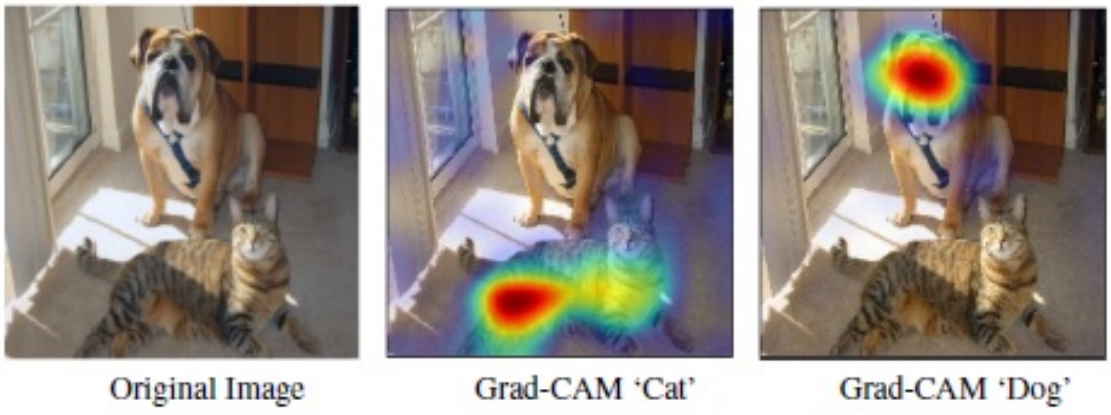
Επίσης, υπάρχουν μέθοδοι ερμηνευσιμότητας σε μεθόδους ταξινόμησης εικόνων. Το Grad-CAM (Gradient-weighted Class Activation Mapping) [39] είναι μια τεχνική που χρησιμοποιείται στο πεδίο της υπολογιστικής όρασης με σκοπό να οπτικοποιηθούν οι περιοχές μίας εικόνας που είναι πιο σημαντικές κατά την πρόβλεψη που γίνεται με ένα συνελικτικό νευρωνικό δίκτυο (CNN). Παρέχει πληροφορίες σχετικά με τη διαδικασία λήψης αποφάσεων των CNN και βοηθά στην κατανόηση ποιών μερών της εικόνας εισόδου συμβάλλουν περισσότερο στην τελική ταξινόμηση.

Όπως ήδη αναπτύχθηκε σε προηγούμενο υποκεφάλαιο, τα συνελικτικά νευρωτικά δίκτυα έχουν την ικανότητα να βρίσκουν μοτίβα μέσα στις εικόνες, αλλά πολλές φορές δεν είναι εύκολο να εξηγηθεί η έξοδος τους. Με απώτερο σκοπό να λυθεί το θέμα αυτό, το Grad-CAM δημιουργεί ένα θερμικό χάρτη (heatmap) που επισημαίνει τις περιοχές της εικόνας που συνεισφέρουν περισσότερο στην έξοδο του δικτύου.

Ο τρόπος με τον οποίο βρίσκει τις περιοχές αυτές είναι χρησιμοποιώντας τις παραγωγούς από το τελευταίο στρώμα νευρώνων του δικτύου και υπολογίζοντας την σημαντικότητα κάθε χάρτη χαρακτηριστικών (feature map) βρίσκοντας τον μέσο όρο των παραγώγων σε σχέση με την κλάση εξόδου. Ο χάρτης χαρακτηριστικών σε ένα CNN αντιπροσωπεύει μια συλλογή φίλτρων που εξάγουν συγκεκριμένα οπτικά μοτίβα ή χαρακτηριστικά από μια εικόνα εισόδου. Έπειτα, οι τιμές σημαντικότητας χρησιμοποιούνται για να δημιουργηθεί ένας σταθμισμένος χάρτης χαρακτηριστικών, ενισχύοντας τις πιο σημαντικές περιοχές που σχετίζονται με την προβλεπόμενη κλάση.

Για τη δημιουργία ενός θερμικού χάρτη Grad-CAM, οι σταθμισμένοι χάρτες χαρακτηριστικών συνδυάζονται χρησιμοποιώντας την τεχνική average pooling, η οποία μειώνει τις διαστάσεις των χαρτών χαρακτηριστικών σε μια ενιαία τιμή ανά κανάλι. Έτσι παράγεται ένας σταθμισμένος συνδυασμός των χαρτών χαρακτηριστικών, όπου κάθε κανάλι αντιπροσωπεύει την ανάλογη κλάση.

Τέλος, ο σταθμισμένος συνδυασμός συνδυάζεται γραμμικά με τους αρχικούς χάρτες χαρακτηριστικών για να ληφθεί ο χάρτης θερμότητας Grad-CAM. Ο χάρτης θερμότητας που προκύπτει τονίζει τις περιοχές της εικόνας που επηρεάζουν έντονα την πρόβλεψη του CNN. Με την επικάλυψη αυτού του χάρτη θερμότητας πάνω στην αρχική εικόνα, γίνεται ευκολότερο να κατανοηθούν οι περιοχές της εικόνας οι οποίες ήταν κρίσιμες για την πρόβλεψη της εκάστοτε κλάσης.



Σχήμα 2.26: Θερμικός χάρτης Grad-CAM στην ανίχνευση σκύλου ή γάτας στην εικόνα

Το Grad-CAM έχει πολλά πλεονεκτήματα, όπως ότι αρχικά δεν απαιτεί να τροποποιηθεί η αρχιτεκτονική του CNN, καθιστώντας ευκολότερη την εφαρμογή του σε προεκπαιδευμένα μοντέλα. Είναι επίσης μια γενική τεχνική που μπορεί να χρησιμοποιηθεί σε διαφορετικές αρχιτεκτονικές. Η οπτικοποίηση που παρέχεται από το Grad-CAM κάνει τα μοντέλα πιο διαφανή και ερμηνεύσιμα και μπορεί να βοηθήσει να βρεθούν σφάλματα στο μοντέλο, να εντοπισθούν προκαταλήψεις και να βελτιωθεί η γενική του απόδοση.

2.8 Αξιολόγηση Μοντέλων Μηχανικής Μάθησης

2.8.1 Μετρικές αξιολόγησης

Για να αξιολογηθεί η απόδοση της εκπαίδευσης του εκάστοτε μοντέλου μηχανικής μάθησης, υπολογίζονται διάφορες μετρικές στα δεδομένα ελέγχου (test data). Παρακάτω αναγράφονται οι μετρικές αξιολόγησης μοντέλων ταξινόμησης (επιβλεπόμενης μηχανικής μάθησης).

- **Πίνακας Σύγχυσης (Confusion matrix):** Ο πίνακας σύγχυσης αποτελεί την βασικότερη μετρική αξιολόγησης, καθώς τα περιεχόμενα του χρησιμοποιούνται για να υπολογιστούν πολλές από τις υπόλοιπες μετρικές. Δείχνει τον αριθμό των πραγματικών και προβλεπόμενων ετικετών για κάθε κατηγορία ετικέτας. Οι τέσσερις κύριες κατηγορίες του είναι:

  - **True Positives (TP):** Οι σωστά θετικές προβλέψεις, δηλαδή ο αριθμός των δειγμάτων που προβλέφθηκαν ως θετικές και ήταν όντως θετικές.

  - **True Negatives (TN):** Οι σωστά αρνητικές προβλέψεις, δηλαδή ο αριθμός των δειγμάτων που προβλέφθηκαν ως αρνητικές και ήταν όντως αρνητικές.

  - **False Positives (FP):** Οι ψευδώς θετικές προβλέψεις, δηλαδή ο αριθμός

των δειγμάτων που προβλέφθηκαν ως θετικές ενώ ήταν αρνητικές.

- **False Negatives (FN):** Οι ψευδώς αρνητικές προβλέψεις, δηλαδή ο αριθμός των δειγμάτων που προβλέφθηκαν ως αρνητικές ενώ ήταν θετικές.

Ο πίνακας σύγχυσης συναντάται με την παρακάτω μορφή:

Πίνακας 2.1: Πίνακας Σύγχυσης - Confusion Matrix

		Predicted Class	
		Positive	**Negative**
True Class	**Positive**	TP	FN
	Negative	FP	TN

Συνήθως ο πίνακας σύγχυσης έχει ως γραμμές τις πραγματικές κλάσεις και ως στήλες τις προβλεπόμενες κλάσεις, αλλά είναι πολύ πιθανό να βρεθεί στην βιβλιογραφία ή στο διαδίκτυο ως ο ανάστροφος του. Δηλαδή, ως οι προβλεπόμενες κλάσεις σε γραμμές και οι πραγματικές κλάσεις σε στήλες.

Στην πραγματικότητα, ένας πίνακας σύγχυσης έχει N x N διαστάσεις, όπου N ο αριθμός των ετικετών. Δηλαδή, για δυαδική ταξινόμηση θα είναι 2 x 2, ενώ για την περίπτωση του προβλήματος της παρούσας διπλωματικής εργασίας, ο πίνακας θα είναι 7 x 7. Οι τέσσερις αριθμοί που αντιστοιχούν στα TP, FP, TN και FN υπολογίζονται ξεχωριστά για κάθε κλάση και μετά υπολογίζονται οι συνολικοί αντίστοιχοι αριθμοί.

- **Ορθότητα (Accuracy):** Η ορθότητα είναι από τις βασικότερες μετρικές που χρησιμοποιούνται για την αξιολόγηση ενός μοντέλου επιβλεπόμενης μάθησης, αν και αρκετά απλή. Μετράει τον αριθμό σωστών προβλέψεων σε σχέση με τον συνολικό αριθμό δειγμάτων.

  Η ορθότητα υπολογίζεται ως εξής:

$$\text{Accuracy} = \frac{\text{Σωστές προβλέψεις}}{\text{Συνολικές προβλέψεις}} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{P + N}$$

  Σημαντικό είναι να αναφέρουμε πως η λέξη "accuracy" μπορεί να μεταφραστεί από την αγγλική γλώσσα ως "ορθότητα" αλλά και ως "ακρίβεια". Επειδή, ωστόσο, η ακρίβεια μπορεί να αντιστοιχεί και στην λέξη precision, η οποία αναφέρεται σε άλλη μετρική αξιολόγησης, καλό είναι οι δύο λέξεις να μην συγχέονται.

- **Ακρίβεια (Precision)**: Αναφέρεται στο ποσοστό των σωστών θετικών προβλέψεων σε σχέση με τον αριθμό των παρατηρήσεων που προβλέφθηκαν ως θετικές. Δίνεται από τον τύπο:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- **Ανάκληση (Recall)**: Ονομάζεται επίσης και ευαισθησία (sensitivity) και True Positive Rate (TPR). Αναφέρεται στο ποσοστό των σωστών θετικών προβλέψεων ως προς τον συνολικό αριθμό των θετικών παρατηρήσεων. Δίνεται από τον τύπο:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **F1-score**: Πρόκειται για τον αρμονικό μέσο ανάκλησης και ακρίβειας. Χρησιμοποιείται ευρέως όταν οι ετικέτες κατανέμονται άνισα (imbalanced dataset), όπως για παράδειγμα όταν έχουμε 100 δείγματα κλάσης A και 50 δείγματα κλάσης B. Εκφράζεται μαθηματικά με τον παρακάτω τύπο:

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

- **Matthew's Correlation Coefficient (MCC):** Εκφράζει την συσχέτιση μεταξύ των προβλεπόμενων και των πραγματικών τιμών, ενώ λαμβάνει υπόψιν και τις αρνητικές όσο και τις θετικές κλάσεις. Χρησιμοποιείται περισσότερο σε προβλήματα δυαδικής ταξινόμησης και δίνεται από τον παρακάτω τύπο:

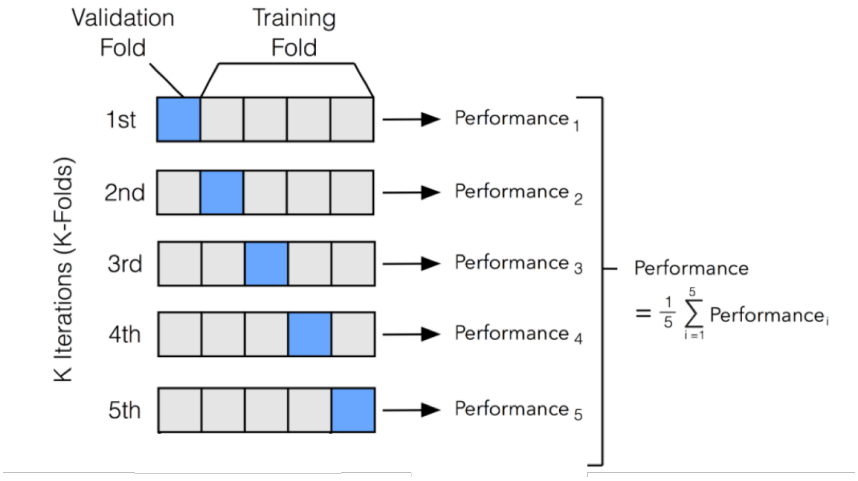
$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

- **Ειδικότητα (Specificity):** Συναντάται επίσης και ως επιλεκτικότητα (selectivity) ή true negative rate (TNR). Εκφράζει το μέρος των δειγμάτων που προβλέφθηκαν αληθώς αρνητικά από όλες τις αρνητικές προβλέψεις. Μεγάλη ειδικότητα σημαίνει πως το μοντέλο αποφεύγει επιτυχώς τους "λάθος συναγερμούς", δηλαδή τις ψευδώς θετικές προβλέψεις.

$$\text{TNR} = \frac{TN}{TN + FP}$$

- **Receiver Operating Characteristic Curve (ROC Curve):** Η καμπύλη ROC αποτελεί γραφικό εργαλείο για την οπτικοποίηση της απόδοσης ενός μοντέλου. Δείχνει την σχέση του True Positive Rate και False Positive Rate σε πολλαπλά επίπεδα κατωφλίου.

  Επίσης, μετριέται και η περιοχή κάτω από την καμπύλη, γνωστή και ως Area Under the Curve (AUC). Όσο πιο μεγάλη είναι η περιοχή κάτω από την κα-



Σχήμα 2.27: Παράδειγμα 5-fold Cross Validation

μπύλη ROC, τόσο καλύτερη είναι και η απόδοση του μοντέλου. Ο αριθμός αυτός καθιστά συγκρίσιμες τις καμπύλες που επικαλύπτονται πολύ στην γραφική παράσταση ειδάλλως θα ήταν δύσκολο να συμπειράνουμε ποια είναι η καλύτερη.

2.8.2 Τεχνικές αξιολόγησης

- **Cross Validation:** Στα ελληνικά είναι γνωστή και ως διασταυρούμενη επιτήρηση. Κατά την διαδικασία αυτή, τα δεδομενα αρχικά χωρίζονται σε k μέρη, εξού και η φράση "k-fold Cross Validation". Έπειτα, ένα από τα κομμάτια διατηρείται στην άκρη ως σετ ελέγχου, ενώ τα υπόλοιπα κομμάτια χρησιμοποιούνται ως σετ εκπαίδευσης. Η διαδικασία αυτή επαναλαμβάνεται μέχρι κάθε ξεχωριστό κομμάτι να έχει χρησιμοποιηθεί ως τεστ ελέγχου, δηλαδή για k επαναλήψεις.

  Ο βασικός στόχος του Cross Validation είναι να προσομοιωθεί η απόδοση ενός μοντέλου σε άγνωστα δεδομένα. Για αυτόν τον λόγο, το μοντέλο εκπαιδεύεται k φορές και συνήθως μετράται ο μέσος όρος της εκάστοτε μετρικής, όπως ο μέσος όρος της ορθότητας των μοντέλων.

- **Leave-One-Out Cross Validation:** Αν n ο αριθμός των παρατηρήσεων του dataset και k=n, τότε έχουμε το λεγόμενο Leave-One-Out Cross Validation, όπου κάθε φορά μόνο μια παρατήρηση κρατάται στην άκρη για έλεγχο. Μια τέτοια διαδικασία μπορεί να είναι κοστοβόρα σε μνήμη και χρόνο, ιδίως όταν υπάρχουν πολλά δείγματα.

2.9 Επιλογή παραμέτρων (Hyperparameter Tuning)

Οι μέθοδοι επιλογής παραμέτρων που θα παρουσιαστούν στην συνέχεια είναι μεθοδοι που βοηθούν στην βελτιστοποίηση του μοντέλου, προσπαθώντας να βρουν τις ιδανικές τιμές των παραμέτρων του.

2.9.1 Αναζήτηση τύπου grid (Grid Search)

Η αναζήτηση τύπου grid, ή Grid Search, είναι μια αναζήτηση η οποία λαμβάνει υπόψιν κάθε πιθανό συνδυασμό παραμέτρων. Το grid search λαμβάνει έναν χώρο παραμέτρων από τον χρήστη, δηλαδή ένα γνωστό "πλέγμα" παραμέτρων, και εκτελεί μια εξαντλητική αναζήτηση προσπαθώντας να βρει το σετ παραμέτρων που αυξάνει μία μετρική, η οποία θα πρέπει επίσης να προκαθοριστεί από τον χρήστη. Για παράδειγμα, σε ένα πρόβλημα ταξινόμησης, η μέθοδος grid search θα αναζητήσει συνήθως τις παραμέτρους που μεγιστοποιούν το accuracy.

2.9.2 Τυχαία αναζήτηση (Random Search)

Η μέθοδος τυχαίας αναζήτησης, ή random search, χρησιμοποιεί τυχαίους συνδυασμούς παραμέτρων με σκοπό να βρει τον καλύτερο συνδυασμό. Καθώς πρόκειται για μια επαναληπτική διαδικασία, σε κάθε επανάληψη λαμβάνονται τυχαίοι αριθμοί και έτσι μπορούμε να οδηγηθούμε ταχύτερα σε ένα ιδανικό αποτέλεσμα, σε αντίθεση με την μέθοδο grid search. Η μέθοδος random search μπορεί, επίσης, να εκτελεστεί παράλληλα πολύ εύκολα, δοκιμάζοντας πολλές παραμέτρους σε παράλληλο χρόνο.

2.9.3 Αναζήτηση Bayes (Bayesian Search)

Η μέθοδος Bayesian Search ή Bayesian Optimization είναι μια μέθοδος αναζήτησης η οποία εκμεταλλεύεται το θεώρημα Bayes για να βελτιστοποιήσει ένα μοντέλο. Η βασική διαφορά μεταξύ αυτής και των άλλων μεθόδων είναι πως η Bayesian Search βελτιστοποιεί την επιλογή παραμέτρων σε κάθε επανάληψη με βάση τις προηγούμενες παραμέτρους. Έτσι, αντί να επιλέγει στην τύχη την επόμενη τιμή μίας παραμέτρου, βελτιστοποιεί την επιλογή και πιθανότατα φτάνει στο καλύτερο σετ παραμέτρων αρκετά γρήγορα. Η μέθοδος Bayesian Search μπορεί να είναι χρήσιμη σε μεγάλο όγκο δεδομένων όπου η διαδικασία μάθησης είναι πιο αργή.

2.10 Συνήθη προβλήματα στην Μηχανική Μάθηση

2.10.1 Προβλήματα των δεδομένων

Ένα πρώτο και βασικό πρόβλημα που αντιμετωπίζεται συχνά στον τομέα της μηχανικής μάθησης είναι η ποιότητα τον δεδομένων. Ουκ ολίγες φορές τα δεδομένα είναι λειψά ή χρειάζονται πολλούς μετασχηματισμούς πριν δωθούν σε κάποιο μοντέλο για εκπαίδευση. Πολύ συχνά όμως, υπάρχει το λεγόμενο imbalanced dataset, όπου έχουμε περισσότερες παρατηρήσεις για μία κλάση, έναντι των άλλων. Αυτό μπορεί

να προκαλέσει προβλήματα στην εκπαίδευση και την αξιολόγηση του μοντέλου, γιατί το μοντέλο μαθαίνει καλύτερα την επικρατούσα κλάση, και ως αποτέλεσμα δεν μπορεί να προβλέψει τις κλάσεις που υπάρχουν σε μικρότερα ποσοστά.

Για να καταπραϋνθεί αυτό το φαινόμενο, υπάρχουν διάφορες τεχνικές. Κάποιες από αυτές είναι η υποδειγματολειψία της κλάσης πλειοψηφίας, η υπερδειγματολειψία της κλάσης μειοψηφίας, η χρήση μετρικών όπως η ευαισθησία ή το F1 score και η χρήση προηγμένων αλγόριθμων που είναι σχεδιασμένοι για τέτοιες περιπτώσεις.

2.10.2 Overfitting και underfitting

Ένα άλλο πρόβλημα που συναντάται συχνά έχει να κάνει με την εκπαίδευση των μοντέλων και ονομάζεται overfitting [2]. Το overfitting είναι ένα φαινόμενο κατά το οποίο το μοντέλο μαθαίνει πολύ καλά τα δεδομένα εκπαίδευσης και δυσκολεύεται να προβλέψει σωστά δεδομένα που δεν του είναι γνωστά. Δηλαδή, το μοντέλο δυσκολεύεται να γενικεύσει έτσι ώστε να μπορεί να προβλέψει όσο το δυνατόν περισσότερες διαφορετικές παρατηρήσεις. Το overfitting οδηγεί σε υποβέλτιστα αποτελέσματα και σε χαμηλή απόδοση.

Τεχνικές καταπολέμησης του overfitting αποτελούν η συλλογή νέων δεδομένων, η μείωση της πολυπλοκότητας του μοντέλου και η χρήση κανονικοποίησης (regularization). Η κανονικοποίηση είναι μια τεχνική που αποτρέπει την υπερεκπαίδευση. Ακόμα, χρησιμοποιείται εξίσου η τεχνική της πρόωρης διακοπής της εκπαίδευσης (early stopping). Η τεχνική αυτή χρησιμοποιείται για να σταματήσει το overfitting χωρίς, όμως, να διακυβεύεται η απόδοση του μοντέλου. Συνήθως, τίθεται από τον χρήστη μια περίοδος "υπομονής" (patience) και παρατηρείται κάποια μετρική στο validation set. Αν, κατά την περίοδο υπομονής, η μετρική αυτή δεν αλλάξει, τότε η εκπαίδευση σταματάει.

Μία ακόμη μέθοδος που επιστρατεύεται για να αποφευχθεί το overfitting και να ενισχυθεί η απόδοση του μοντέλου είναι η μέθοδος dropout [40]. Το dropout εφαρμόζεται στα μοντέλα βαθιάς μάθησης και ουσιαστικά "απενεργοποιεί" προσωρινά ένα ποσοστό νευρώνων κατά τη διάρκεια της εκπαίδευσης. Αυτή η προκαλούμενη απόρριψη νευρώνων εισάγει μια μορφή τυχαιότητας που εμποδίζει το δίκτυο να βασιστεί υπερβολικά σε συγκεκριμένες συνδέσεις ή χαρακτηριστικά, αυξάνοντας έτσι την ανθεκτικότητα και την προσαρμοστικότητά του. Ως αποτέλεσμα, μειώνεται ο κίνδυνος υπερπροσαρμογής ενισχύοντας το δίκτυο και καθιστώντας το πιο ανθεκτικό και ικανό να γενικεύσει σε διάφορα δεδομένα εισόδου.

Επιπροσθέτως, αντίστοιχα με το overfitting, υπάρχει και το πρόβλημα του underfitting. Το φαινόμενο του underfitting συναντάται όταν το μοντέλο αδυνατεί να προσαρμοστεί στο training set και, κατά συνέπεια, έχει χαμηλή απόδοση στο test set. Οι αιτίες του underfitting ποικίλλουν ανά περίπτωση, όμως συνήθως underfit-

[2] Στα ελληνικά το overfitting συναντάται συχνά ως "υπερπροσαρμογή". Ωστόσο, ο αγγλικός όρος φαίνεται να είναι επικρατέστερος.

ting συναντάται όταν το μοντέλο δεν εκπαιδεύτηκε αρκετά ή εκπαιδεύτηκε με πολύ λίγα δεδομένα. Παρολαυτά, συνήθως βελτιώνεται αμέσως είτε μεγαλώνοντας το σετ εκπαίδευσης, είτε αλλάζοντας τις παραμέτρους του μοντέλου, είτε προσθέτοντας επιπλέον δείγματα.

2.10.3 Το φαινόμενο Plateau

Το φαινόμενο Plateau αναφέρεται στο γεγονός ότι μετά από κάποιο συγκεκριμένο αριθμό βημάτων εκπαίδευσης το μοντέλο παύει να βελτιώνεται και η συνάρτηση κόστους μένει σταθερή ή μειώνεται με πολύ αργούς ρυθμούς. Το φαινόμενο αυτό παρατηρείται πολύ συχνά στα Τεχνητά Νευρωνικά Δίκτυα κατά την εκπαίδευσή τους και, αν δεν ληφθεί δράση, οδηγεί σε υποβέλτιστα αποτελέσματα.

Μία λύση στο πρόβλημα αυτό είναι η μέθοδος (ή το ¨callback¨) ReduceLROnPlateau. Η μέθοδος αυτή μειώνει το learning rate σταδιακά με βάση έναν δοσμένο παράγοντα (π.χ. κατά 0.001 κάθε φορά), όταν η διαδικασία της μάθησης φαίνεται να έχει κολλήσει σε ένα συγκεκριμένο σημείο. Το learning rate μειώνεται μόνο αν μείνει σταθερό για ένα δοσμένο αριθμό εποχών, γνωστό και ως patience.

2.10.4 Πρόβλημα των εξαφανιζόμενων παραγώγων (vanishing gradients)

Το πρόβλημα των εξαφανιζόμενων κλίσεων, ή αλλιώς vanishing gradients, είναι ένα θέμα που προκύπτει κατά τη διάρκεια εκπαίδευσης τεχνητών νευρωνικών δικτύων και ιδίως των επαναλαμβανόμενων νευρωνικών δικτύων (RNN). Η πρόκληση αυτή εμφανίζεται όταν οι παράγωγοι της συνάρτησης κόστους γίνονται πολύ μικρές καθώς διαδίδονται προς τα πίσω με την τεχνική της οπισθοδιάδοσης (backpropagation).

Κατά τη διάρκεια της οπισθοδιάδοσης, οι παράγωγοι χρησιμοποιούνται για την ενημέρωση των βαρών μέσω του αλγορίθμου gradient descent. Ωστόσο, όταν οι κλίσεις μικραίνουν κατά πολύ, οι ενημερώσεις στις παραμέτρους γίνονται επίσης πολύ μικρές και έτσι η διαδικασία της εκπαίδευσης επιβραδύνεται ή μένει στάσιμη. Το πρόβλημα των vanishing gradients είναι ιδιαίτερα έντονο σε ΤΝΔ με πολλά κρυφά στρώματα, αφού οι παράγωγοι πολλαπλασιάζονται κατά μήκος του δικτύου και μπορούν να μικρύνουν υπερβολικά, ειδικά όταν χρησιμοποιούνται συναρτήσεις ενεργοποίησης όπως η σιγμοειδής ή η υπερβολική εφαπτομένη, οι οποίες προκαλούν κορεσμό σε ακραίες τιμές.

Το πρόβλημα αυτό μπορεί να μετριαστεί με την χρήση συναρτήσεων ενεργοποίησης, όπως η ReLU, οι οποία δεν προκαλεί κορεσμό στις ακραίες τιμές, ή με την τεχνική batch normalization, η οποία κανονικοποιεί το κάθε batch [41].

2.10.5 Πρόβλημα της απότομης αύξησης των παραγώγων (exploding gradients)

Το πρόβλημα αυτό προκύπτει όταν υπάρχει η αντίθετη τάση στις παραγώγους, δηλαδή όταν γίνονται πολύ μεγάλες. Αυτό μπορεί να οδηγήσει σε ασταθή μάθηση ή ακόμα και σε απόκλιση του δικτύου από τον πραγματικό στόχο της εκπαίδευσης. Μια συνήθης τακτική που εφαρμόζεται για να καταπραϋνθεί αυτό το πρόβλημα είναι αυτή του gradient clipping. Ουσιαστικά, οι παράγωγοι "ψαλιδίζονται" με βάση ένα συγκεκριμένο κατώφλι κατά τη διάρκεια της οπισθοδιάδοσης. Ακόμα, η τεχνική batch normalization είναι επίσης μια λύση τόσο για το πρόβλημα των exploding όσο και των vanishing gradients.

Κεφάλαιο 3

Συναφής Βιβλιογραφία

3.1 Εισαγωγή

Το κεφάλαιο αυτό αποτελεί μια σημαντική ενότητα που εξετάζει την προηγούμενη έρευνα και τις σχετικές εργασίες που έχουν διεξαχθεί στον ευρύτερο τομέα της αναγνώρισης συναισθημάτων σε εικόνες προσώπων, χρησιμοποιώντας τεχνητά νευρωνικά δίκτυα. Θα εξετασθούν προηγούμενες μελέτες και τεχνικές που έχουν αναπτυχθεί για την αναγνώριση συναισθημάτων σε εικόνες προσώπων. Επίσης, θα εξετασθεί η χρήση τεχνικών βαθιάς μάθησης, όπως τα τεχνητά νευρωνικά δίκτυα. Θα αναλυθούν διάφορες αρχιτεκτονικές νευρωνικών δικτύων που έχουν προταθεί, όπως τα συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNNs).

Θα αναλυθούν επίσης οι μέθοδοι προεπεξεργασίας εικόνων που χρησιμοποιούνται για την εξαγωγή χαρακτηριστικών από τις εικόνες προσώπων. Αυτές οι μέθοδοι περιλαμβάνουν την απομάκρυνση θορύβου, την κανονικοποίηση, τον προσδιορισμό σημείων αναφοράς και τη μείωση της διάστασης των χαρακτηριστικών.

Τέλος, θα εξετασθεί η απόδοση καθώς και οι προκλήσεις που συναντώνται στην αναγνώριση συναισθημάτων σε εικόνες προσώπων. Θα εξετάσουμε την ορθότητα και την απόδοση των μεθόδων, καθώς και τις προκλήσεις που προκύπτουν από την ποικιλία των συναισθηματικών εκφράσεων, τις αναλογίες και την επιρροή του φωτισμού και της γωνίας λήψης.

Αυτό το κεφάλαιο παρέχει μια επισκόπηση της προηγούμενης έρευνας στην αναγνώριση συναισθημάτων σε εικόνες προσώπων και αποτελεί ένα θεμελιώδες μέρος της εργασίας, προσφέροντας την απαραίτητη βάση γνώσης για την περαιτέρω ανάπτυξη και αξιολόγηση της προτεινόμενης μεθοδολογίας για την αναγνώριση συναισθημάτων στο πλαίσιο της διπλωματικής εργασίας.

3.2 Υπάρχουσες μέθοδοι βαθιάς μάθησης

Οι Nwosu et al. [42] ανέπτυξαν ένα τεχνητό νευρωνικό δίκτυο (CNN) δύο καναλιών το οποίο προβλέπει με μεγάλη ορθότητα τα συναισθήματα από εικόνες προσώπου. Το πρώτο βήμα ήταν η προεπεξεργασία της εικόνας, η οποία σε πρώτο στάδιο συμπεριλάμβανε την ανίχνευση προσώπου με τον αλγόριθμο Viola-Jones [43] και την κανονικοποίηση της εικόνας. Το δεύτερο βήμα ήταν η απομόνωση χαρακτηριστικών, και συγκερκιμένα του στόματος και των ματιών. Έπειτα, τα δύο μέρη αυτά, εισήχθησαν το καθένα σε ένα CNN. Τα δύο CNN ενώθηκαν αργότερα με ένα πλήρως συνδεδεμένο στρώμα νευρώνων (Fully Connected Layer), το οποίο προέβλεπε τα επτά συναισθήματα. Η παραπάνω μέθοδος εφαρμόσθηκε σε δύο datasets, το JAFFE [44] [45], και στο CK+, στα οποία πέτυχαν 97.71 % και 95.71 % αντίστοιχα.

Οι Khandait et al. [46] πρότειναν μια μέθοδο με την οποία προβλέπουν τα συναισθήματα σε μια εικόνα μέσω ενός απλού πλήρως συνδεδεμένου τεχνητού νευρωνικού δικτύου. Αρχικά, προεπεξεργάστηκαν τις εικόνες με τον εξής τρόπο: πρώτα απομόνωσαν το πρόσωπο με μια μέθοδο κατά την οποία η περιοχή του προσώπου τμηματοποιείται χρησιμοποιώντας μορφολογικές λειτουργίες επεξεργασίας εικόνας (διαστολή, ανακατασκευή με διάβρωση, συμπλήρωση κ.α.). Έπειτα, απομόνωσαν τα χαρακτηριστικά του προσώπου και κατασκευάζουν διανύσματα που τα περιγράφουν. Τέλος, εκπαίδευσαν ένα νευρωνικό δίκτυο που προβλέπει τα 7 συναισθήματα με βάση τα διανύσματα αυτά. Η μέθοδος αυτή απέφερε 95.26 % ορθότητα στο test set των εικόνων JAFFE [44] [45].

Οι Yolcu et al. [47] πρότειναν μια αρχιτεκτονική CNN η οποία πρώτα απομονώνει τα χαρακτηριστικά του προσώπου και φτιάχνει μία "προσωρινή" εικόνα από αυτά. Έπειτα, συνδυάζουν την κανονική φωτογραφία με την προσωρινή και προβλέπουν το συναίσθημα με ένα άλλο CNN καθάυτόν τον τρόπο. Έτσι, συνδυάζουν την γενική πληροφορία που έρχεται από ολόκληρο το πρόσωπο με τα τοπικά χαρκακτηριστικά του προσώπου, κάνοντας έτσι το μοντέλο πιο δυνατό. Η παραπάνω τεχνική εφαρμόσθηκε στις εικόνες που προήλθαν από το Radboud Face Database [48] και απέφερε 94.44 % ορθότητα.

Οι Arora et al. [49] χρησιμοποίησαν επίσης ένα CNN για να εντοπίσουν επτά συναισθήματα. Αρχικά, προεπεξεργάστηκαν την εικόνα, χρησιμοποιώντας διάφορα φίλτρα (όπως Γκαουσιανό φίλτρο) για να αφαιρέσουν τον θόρυβό της. Έπειτα αφαίρεσαν το φόντο της εικόνας και δημιούργησαν ένα διάνυσμα που περιγράφει την εικόνα, με την χρήση ενός CNN το οποίο εντόπισε τα κύρια σημεία του προσώπου. Για να υπολογισθεί αυτό το διάνυσμα, χρησιμοποίησαν τεχνικές που βασίστηκαν στην γεωμετρία και σε ειδικά σημεία του προσώπου, αλλά και μια ολιστική τεχνική που βασίστηκε σε όλο το πρόσωπο. Επίσης, χρησιμοποίησαν μια τεχνική που λάμβανε υπόψιν την κατανομή των χρωμάτων στην εικόνα. Στο τέλος, η πρόβλεψη έλαβε χώρα μέσω των τελευταίων στρωμάτων του CNN. Η μέθοδος αυτή απέφερε 97 % ορθότητα στο test set εικόνων, το οποίο προήλθε από ένα άγνωστο σετ εικόνων που δεν αναφέρεται, καθώς και από το FER2013 [50].

Οι Lopes et al. [51] επικεντρώθηκαν στις περιπτώσεις όπου οι διαθέσιμες εικόνες είναι λίγες. Αρχικά, παρήγαγαν νέες εικόνες με βάση αυτό το μικρό σετ εικόνων, εφαρμόζοντας διάφορες μεταμορφώσεις στην κάθε εικόνα, όπως περιστροφή, καθρεπτισμό και διαστρέβλωση. Έπειτα, διόρθωσαν τον προσανατολισμο, έτσι ώστε το πρόσωπο να έρθει παράλληλα με τον οριζόντιο άξονα της εικόνας, εμπλουτίζοντας έτσι την γεωμετρική ακριβεία της κάθε εικόνας. Το επόμενο βήμα ήταν η περικοπή της εικόνας γύρω από το πρόσωπο, για να αποφευχθεί ο περιττός θόρυβος γύρω από αυτό. Ακόμη, εφαρμόσθηκε μείωση των διαστάσεων των εικόνων και κανονικοποίηση, τα οποία θα καθιστούσαν την μετέπειτα διαδικασία συνέλιξης πιο εύκολη. Στο τέλος της προεπεξεργασίας, εφαρμόσθηκε κανονικοποίηση της έντασης των εικονοστοιχείων, για να εξομαλυνθεί η φωτεινότητα και η αντίθεση μεταξύ των εικόνων.

Ο πυρήνας της μεθόδου αυτής είναι ένα CNN, το οποίο αποτελείται από δύο συνελικτικά στρώματα, δύο στρώματα υποδειγματολειψίας και ένα πλήρως συνδεδεμένο στρώμα. Το CNN αυτό μαθαίνει να εξάγει στοιχειώδη οπτικά χαρακτηριστικά και στοιχεία που σχετίζονται με τις εκφράσεις του προσώπου αφού εκπαιδεύθηκε με τον αλγόριθμο Stochastic Gradient Descent. Η συγκεκριμένη μέθοδος πέτυχε ορθότητα της τάξης του 96.76 % στο σετ εικόνων προσώπου Extended Cohn-Kanade (CK+) [52] και 86.74 % ορθότητα στο JAFFE.

Ο Ninad Mehendale [53] πρότεινε μια μέθοδο με όνομα Facial Emotion Recognition using Convolutional Neural Networks (FERC) η οποία χρησιμοποιεί ένα CNN δύο επιπέδων. Το πρώτο επίπεδο περιλαμβάνει την αφαίρεση φόντου. Το δεύτερο επίπεδο εστιάζει στην εξόρυξη χαρακτηριστικών του προσώπου με την χρήση ενός ακόμα CNN. Τα χαρακτηριστικά του προσώπου αναπαριστώνται με την βοήθεια ενός διανύσματος (το λεγόμενο expressional vector - EV) το οποίο υπολογίσθηκε με την βοήθεια ενός perceptron το οποίο εφαρμόσθηκε στην εικόνα χωρίς το φόντο. Το διάνυσμα αυτό καταγράφει αλλαγές στην έκφραση και αποτελείται από 24 τιμές που αντιπροσωπεύουν κανονικοποιημένες Ευκλείδειες αποστάσεις μεταξύ διαφορετικών μερών του προσώπου. Η αρχιτεκτονική του CNN χρησιμοποιεί συνελικτικά στρώματα με φίλτρα για ανίχνευση προτύπων και το τελευταίο στρώμα είναι ένα perceptron που βελτιστοποιεί τις τιμές συντελεστή κλίμακας και εκθέτη.

Στο άρθρο αυτό ο Mehendale συζητά επίσης και την εξαγωγή καρέ από βίντεο με σκοπό την πρόβλεψη του συναισθήματος. Το πλάνο με το μέγιστο συγκεντρωτικό άθροισμα λευκών εικονοστοιχείων, που λαμβάνεται μέσω της ανίχνευσης άκρων Canny [54], επιλέγεται ως είσοδος για το FERC, και έπειτα εφαρμόζεται η αφαίρεση φόντου. Μετέπειτα, η εικόνα χωρίζεται σε επικαλυπτόμενους πίνακες μεγέθους 3×3 και εφαρμόζονται συνελικτικά φίλτρα για την εξαγωγή χαρακτηριστικών. Η συγκεκριμένη μέθοδος απέφερε έως και 96 % ορθότητα σε ένα σετ εικόνων, το οποίο δυστυχώς δεν διευκρινίζεται.

Οι Mollahosseini et al. [55] πρότειναν μια αρχιτεκτονική ΤΝΔ η οποία χρησιμοποιεί ειδικά στρώματα νευρώνων, τα λεγόμενα inception layers, καθένα από τα οποία λειτουργεί ως ένα νευρωνικό μικροδίκτυο το οποίο αποτελεί δομικό στοιχείο του με-

γαλύτερου δικτύου. Το inception layer έγινε γνωστό από ερευνητές της Google από την αρχιτεκτονική Inception, η οποία είναι η ραχοκοκαλιά του διάσημου μοντέλου GoogLeNet [56]. Η βασική ιδέα πίσω από το στρώμα αυτό είναι η χρήση πολλαπλών μεγεθών φίλτρων και λειτουργιών pooling παράλληλα, με σκοπό τον εντοπισμό χαρακτηριστικών σε διαφορετικές χωρικές κλίμακες μέσα στο ίδιο επίπεδο. Με αυτόν τον τρόπο, το δίκτυο μπορεί να εξάγει χαρακτηριστικά σε διάφορα επίπεδα αφαίρεσης και να συλλάβει διαφορετικά μοτίβα στα δεδομένα εισόδου.

Το προτεινόμενο δίκτυο αποτελείται από δύο συνελικτικά στρώματα τα οποία ακολουθούνται από ενα στρώμα που εκτελεί την λειτουργία max pooling και έπειτα από τέσσερα στρώματα inception. Το μοντέλο αυτό επέφερε πολύ ικανοποιητικά αποτελέσματα στα περισσότερα σετ εικόνων που εφαρμόσθηκε. Δηλαδή: 94.7 % στο σετ εικόνων MultiPIE [57], 77.6 % στο MMI [58], 55 % στο DISFA [59], 76.7 % στο FERA [60], 47.7 % στο SFEW [61], 93.2 % στο CK+ [52], 66.4 % στο FER2013 [50].

Οι Oztel et al. [62] χρησιμοποίησαν την τεχνική transfer learning για να εκμεταλλευτούν την δύναμη ήδη εκπαιδευμένων μοντέλων με σκοπό την πρόβλεψη των συναισθημάτων. Πιο συγκεκριμένα, οι Oztel et al. χρησιμοποίησαν τα δίκτυα VGG [21] και AlexNet [22] κάνοντας δύο πειράματα με το καθένα: το πρώτο ήταν να εκπαιδεύσουν το κάθε δίκτυο από την αρχή και το δεύτερο να χρησιμοποιήσουν transfer learning. Συνολικά, δηλαδή, εκπαίδευσαν τέσσερα μοντέλα. Το μοντέλο VGG με την τεχνική transfer learning απέφερε την καλύτερη ορθότητα (98.33 %) στο σετ εικόνων RaFD [48].

Οι Palaniswamy [63] δημοσίευσαν την μέθοδο DPIIER (Deep learning Pose Illumination Invariant Emotion Recognition) η οποία μπορεί να ταξινομήσει πέντε βασικά συναισθήματα (θυμό, χαρά, έκπληξη, αηδία και το ουδέτερο συναίσθημα). Οι Palaniswamy πρότειναν μια αρχιτεκτονική CNN 15 στρωμάτων, η οποία αποτελείται από τρία συνελικτικά στρώματα, όπου το καθένα ακολουθείται από συνάρτηση ενεργοποίησης ReLU και στρώματα pooling, και από πλήρως συνδεδεμένα στρώματα με συνάρτηση ενεργοποίησης Softmax. Χρησιμοποίησαν επίσης 3-fold Cross Validation και βρήκαν τις βέλτιστες παραμέτρους για το δίκτυο. Παρόλο που εφάρμοσαν το μοντέλο και στα σετ εικόνων KDEF [64], JAFFE [44] [45] και CK+ [52], τελικώς εκπαίδευσαν το μοντέλο στο σετ εικόνων Multi-PIE [57], καθώς περιείχε εικόνες με διαφορετικές γωνίες λήψεις και φωτισμό, κάνοντας έτσι το μοντέλο πιο ισχυρό. Η μέση ορθότητα που πέτυχαν, χρησιμοποιώντας Cross Validation, ήταν 96.55 % στην βάση εικόνων Multi-PIE [57].

Οι Munsif et al. [65] χρησιμοποίησαν μεθόδους βαθιάς μάθησης για να παρακολουθήσουν νευρολογικές διαταραχές όπως το Alzheimer, το Parkinson και τα εγκεφαλικά επεισόδια, μέσω της έκφρασης του προσώπου του ασθενούς. Ένα από τα προτερήματα του μοντέλου αυτού είναι πως είναι αρκετά "ελαφρύ" και μπορεί να τρέξει σε κινητές συσκευές. Αρχικά, εντόπισαν και απομόνωσαν το πρόσωπο του ασθενούς, με τον γνωστό αλγόριθμο Viola-Jones [43], αφού μετέτρεψαν την εικόνα σε ασπρόμαυρη. Έπειτα μίκρυναν τις εικόνες σε μέγεθος 148×148 pixel.

Εν συνεχεία, έδωσαν ως είσοδο τις εικόνες σε ένα CNN το οποίο αποτελείται από τα εξής μέρη: έξι συνελικτικά στρώματα με διάφορα μεγέθη φίλτρων (3×3 στα πρώτα τέσσερα στρώματα και 5×5 στα επόμενα δύο) με συνάρτηση ενεργοποίησης ReLU, πέντε στρώματα pooling μετά από κάθε συνελικτικό στρώμα εκτός από το πρώτο, επτά στρώματα κανονικοποίησης (batch normalization layers - BNLs) μετά από κάθε συνελικτικό στρώμα, και τέλος τρία πλήρως συνδεδεμένα στρώματα, εκ των οποίων το ένα είναι η έξοδος. Το συγκεκριμένο μοντέλο προβλέπει μόνο τέσσερα συναισθήματα (χαρά, λύπη, θυμό και το ουδέτερο συναίσθημα). Το μοντέλο αυτό απέφερε μέγιστη ορθότητα 97 % στο σετ εικόνων KDEF [64].

3.3 Άλλες μέθοδοι

Οι μέθοδοι που θα αναφερθούν σε αυτό το υποκεφάλαιο δεν περιορίζονται αποκλειστικά σε συνελικτικά νευρωνικά δίκτυα (CNN) και νευρωνικά, αλλά περιλαμβάνουν και άλλες τεχνικές όπως η ανάλυση κυρίαρχων συνιστωσών (PCA), μέθοδοι φίλτρων και ορισμένες προσαρμοστικές τεχνικές.

Οι Agrawal και Khatri [66] πρότειναν μια μέθοδο που βασίζεται στην ανάλυση κύριων συνιστωσών (Principal Component Analysis - PCA) για να προβλέψουν έξι συναισθήματα. Καταρχάς, προεπεξεργάζονται την εικόνα με τα εξής βήματα: πρώτα ανιχνεύουν το πρόσωπο με τον αλγόριθμο Viola Jones [43], έπειτα εντοπίζουν τα κύρια χαρακτηριστικά του προσώπου (μάτια, μύτη, στόμα, φρύδια και πηγούνι), μετά ανιχνεύουν το χρώμα του δέρματος με σκοπό να την φωτίσουν αν χρειάζεται, αφαιρούν τον θόρυβο χρησιμοποιώντας αλγόριθμους αφαίρεσης θορύβου και τέλος ανιχνεύουν τις ακμές της εικόνας. Έπειτα, εφαρμόζουν ανάλυση κύριων συνιστωσών (PCA), υπολογίζοντας τις ιδιοτιμές και τα ιδιοδιανύσματα της κάθε εικόνας, και μετρούν την ευκλείδια απόσταση από το ουδέτερο συναίσθημα. Η μέθοδος αυτή απέφερε 99,0744 % ορθότητα (99.84 % ορθότητα για εικόνες με πολλαπλά πρόσωπα). Το σετ εικόνων που χρησιμοποιήθηκε δεν αναφέρεται, όμως σύμφωνα με τις εικόνες της δημοσίευσης, πρόκειται για εικόνες των συγγραφέων.

Οι Boughida et al. [67] πρότειναν μια μέθοδο η οποία βασίζεται στα φίλτρα Gabor και σε γενετικούς αλγόριθμους. Ως πρώτο βήμα, βρήκαν τα 68 σημεία αναφοράς (facial landmarks) τα οποία ορίζουν τις περιοχές ενδιαφέροντος (Regions Of Interest - ROI) με την μέθοδο των Sullivan και Kazemi [68]. Οι περιοχές αυτές είναι τα μάτια, τα φρύδια και το στόμα και απομονώθηκαν για να περάσουν στην συνέχεια φίλτρα Gabor 2 συχνοτήτων και 5 προσανατολισμών. Τα φίλτρα Gabor είναι μαθηματικές συναρτήσεις που χρησιμοποιούνται στην επεξεργασία και την ανάλυση υφής της εικόνας. Πήραν το όνομα τους από τον φυσικό Dennis Gabor, ο οποίος τα εισήγαγε στο πλαίσιο της ανάλυσης σημάτων και της θεωρίας της επικοινωνίας. Τα φίλτρα Gabor έχουν σχεδιαστεί για να μιμούνται ορισμένες ιδιότητες της ανθρώπινης οπτικής αντίληψης, ιδιαίτερα την ικανότητα του οπτικού συστήματος να αναλύει υφές και να αναγνωρίζει μοτίβα σε διαφορετικούς προσανατολισμούς και κλίμακες.

Ένα φίλτρο Gabor δίνεται από τον παρακάτω τυπο:

$$\psi(x, y, \overline{\omega}, \partial) = \frac{1}{2\pi\sigma^2} e^{(\frac{x'^2+y'^2}{2\sigma^2})} [e^{i\overline{\omega}x'} - e^{-\frac{\overline{\omega}^2\sigma^2}{2}}]$$

$$x' = xcos\partial + ysin\partial$$

και

$$y' = -xsin\partial + ycos\partial$$

όπου (x, y) είναι η τοποθεσία του εικονοστοιχείου, ω η κεντρική ακτινική συχνότητα, ∂ η κατεύθυνση του φίλτρου Gabor και σ η τυπική απόκλιση του φίλτρου.

Η αναπαράσταση ενός φίλτρου Gabor μίας εικόνας $I(x, y)$ είναι η συνέλιξη της με το φίλτρο, δηλαδή:

$$O(x, y) = I(x, y) * \psi(x, y, \overline{\omega}, \partial)$$

Μετέπειτα, τα διανύσματα που προέκυψαν από τα φίλτρα Gabor συμπτήχθηκαν σε ένα διάνυσμα για κάθε εικόνα, και στην συνέχεια χρησιμοποιήθηκαν για την εκπαίδευση ενός μοντέλου ταξινόμησης SVM. Ύστερα, οι παράμετροι του μοντέλου βελτιστοποιήθηκαν με την χρήση γενετικών αλγόριθμων. Ένας γενετικός αλγόριθμος είναι μια εξελικτική μέθοδος αναζήτησης και βελτιστοποίησης που εμπνέεται από τη φυσική διαδικασία της εξέλιξης και της φυσικής επιλογής. Ο αλγόριθμος αυτός χρησιμοποιείται για την επίλυση προβλημάτων βελτιστοποίησης, όπου πρέπει να βρεθεί η βέλτιστη λύση σε ένα χώρο αναζήτησης, καθώς και για προβλήματα μηχανικής μάθησης και τεχνητής νοημοσύνης. Η μέθοδος αυτή επέφερε ορθότητα 96.3 % στο σετ εικόνων JAFFE [44] [45], 94.2 % στο σετ εικόνων CK [69] και 94.26 % στο σετ εικόνων CK+ [52].

Οι Mehta και Jadhav [70] πρότειναν μια παρόμοια μέθοδο με την προηγούμενη, η οποία χρησιμοποιεί φίλτρα Log Gabor, δηλαδή φίλτρα Gabor σε λογαριθμική κλίμακα. Το πρώτο τους βήμα ήταν να προεπεξεργαστούν την εικόνα, προσαρμόζοντας την αντίθεση και την φωτεινότητα και κανονικοποιώντας την εικόνα. Έπειτα, απομόνωσαν το πρόσωπο και εφάρμοσαν τα φίλτρα Gabor 5 μεγεθών και 8 κατευθύνσεων (5×8) τα οποία κατέληξαν σε 40 εικόνες για κάθε πρόσωπο. Μετέπειτα χρησιμοποιήθηκε ανάλυση κύριων συνιστωσών (Principal Component Analysis - PCA) για να συμπιεστούν οι 40 αυτές εκόνες, είτε συμπιέζοντας προς τις 8 κατευθύνσεις, είτε προς τα 5 μεγέθη. Ως τελευταίο βήμα, έγινε η ταξινόμηση των συναισθημάτων των εικόνων του test set με βάση την ευκλείδια απόστασή των φίλτρων Gabor από τα φίλτρα Gabor του training set. Η μέθοδος αυτή επέφερε 93.57 % ορθότητα στις εικόνες του test set.

Οι Ahmet et al. [71] επιχείρησαν να προβλέψουν τα επτά συναισθήματα χρησιμοποιώντας Local Binary Patterns (LBP) [72]. Η μέθοδος LBP προτάθηκε πρώτα για ανάλυση υφής και έπειτα για αναγνώριση προσώπων και συναισθήματος. Η διαδικασία λειτουργεί ως εξής: για κάθε εικονοστοιχείο στην εικόνα, δημιουργείται ένας δυαδικός κώδικας συγκρίνοντας την τιμή φωτεινότητας του κεντρικού εικονοστοιχείου με αυτήν των γειτόνων του. Έπειτα, οι τιμές των γειτονικών εικονοστοιχείων

συγκρίνονται με την τιμή του κεντρικού εικονοστοιχείου, και αν το γειτονικό εικονοστοιχείο είναι μεγαλύτερο ή ίσο με το κεντρικό, του αντιστοιχεί δυαδική τιμή 1, αλλιώς του αντιστοιχεί δυαδική τιμή 0. Εν συνεχεία, οι δυαδικές τιμές που προκύπτουν από τις συγκρίσεις των γειτονικών εικονοστοιχείων συνενώνονται για να σχηματίσουν έναν LBP κώδικα για το κεντρικό εικονοστοιχείο. Στο τέλος, οι κώδικες LBP συλλέγονται από διάφορες περιοχές της εικόνας, και δημιουργείται ένα ιστόγραμμα που αντιπροσωπεύει την κατανομή των LBP προτύπων σε αυτήν την περιοχή.

Αφού παρήγαγαν τα ιστογράμματα για κάθε εικόνα, οι Ahmet et al. χρησιμοποίησαν ένα μοντέλο ταξινόμησης SVM. Μετά από πολλαπλά πειράματα, η καλύτερη ορθότητα που επετεύχθη ήταν 94.4 % στο σετ εικόνων Cohn-Kanade [69].

Οι Lakshmi και Ponnusamy [73] συνδύασαν την ιδέα των Local Binary Patterns (LBP), των Histogram of Oriented Gradients (HOG) [74] και των autoencoders και έφτιαξαν ένα μοντέλο που προβλέπει με μεγάλη ορθότητα τα συναισθήματα από εικόνες προσώπου. Αρχικά, εντόπισαν και απομόνωσαν το πρόσωπο με την βοήθεια του γνωστού αλγόριθμου Viola Jones [43]. Έπειτα, εφάρμοσαν ένα υψιπερατό φίλτρο Butterworth για να τονισθούν τα στοιχεία του προσώπου, και εντοπίσθηκαν τα μάτια, το στόμα και η μύτη πάλι με την βοήθεια του Viola Jones [43]. Το επόμενο βήμα ήταν να εφαρμόσουν την μέθοδο HOG και την LBP, να ενώσουν τα δύο διανύσματα που υπολογίσθηκαν από τις δύο μεθόδους και να τα συμπιέσουν χρησιμοποιώντας μια σειρά από autoencoders. Το τελευταίο βήμα ήταν να χρησιμοποιήσουν τα συμπιεσμένα διανύσματα ως είσοδο για την εκπαίδευση ενός μοντέλου ταξινόμησης SVM. Τα αποτελέσματα ήταν 97.66 % ορθότητα στο σετ εικόνων CK [69] και 97.67 % στο σετ JAFFE [44] [45].

Οι Deng et al. [75] παρουσίασαν μια μέθοδο η οποία χρησιμοποιεί επίσης φίλτρα Gabor, σε συνδυασμό με Ανάλυση Κύριων Συνιστωσών (Principal Component Analysis - PCA) και Γραμμιικη Διακριτική Ανάλυση (Linear Discriminant Analysis - LDA). Το πρώτο βήμα ήταν να προεπεξεργαστούν την εικόνα, κανονικοποιώντας την φωτεινότητα και προσαρμόζοντας το σχήμα και το μέγεθος. Πιο αναλυτικά, ανίχνευσαν τα σημεία των χαρακτηριστικών του προσώπου (μάτια, μύτη και στόμα), περιέστρεψαν την κάθε εικόνα ώστε να ευθυγραμμιστούν όλες και εντόπισαν το τετράγωνο εντός του οποίου περιέχεται το κάθε πρόσωπο. Έπειτα, προσάρμοσαν το μέγεθος της εικόνας, φροντίζοντας το κέντρο των ματιών να βρίσκεται σε σταθερή θέση. Τέλος, εφάρμοσαν μια μέθοδο εξισορρόπησης ιστογράμματος για να εξισορροπήσουν τον φωτισμό της κάθε εικόνας.

Όσον αφορά το μοντέλο πρόβλεψης, αρχικά χρησιμοποίησαν μια "τράπεζα" από φίλτρα Gabor για να δημιουργήσουν ένα διάνυσμα που περιγράφει την εικόνα. Έπειτα, συμπίεσαν τα διανύσματα χρησιμοποιώντας μόνο PCA, αλλά και συνδυαστικά με την μέθοδο LDA. Στο τέλος, ταξινόμησαν τις εικόνες με βάση την ευκλείδια απόσταση. Η συγκεκριμένη μέθοδος απέφερε 97.33 % χρησιμοποιώντας PCA και LDA συνδυαστικά για την συμπίεση των διανυσμάτων στο σετ εικόνων JAFFE [44] [45].

Οι De et al. [76] πρότειναν μια μέθοδο η οποία χρησιμοποιεί τα λεγόμενα eigenfaces [77] για την ταξινόμηση των διαφόρων συναισθημάτων. Η μέθοδος των eigenfaces πήρε το όνομα της από τα ιδιοδιανύσματα (eigenvectors) τα οποία χρησιμοποιούνται κατά τον υπολογισμούς της ανάλυσης κυρίως συνιστωσών (PCA). Ουσιαστικά, εφαρμόζεται PCA στο σετ εικόνων και εξάγονται οι κύριες συνιστώσες τους. Οι κύριες συνιστώσες, οι οποίες αναπαριστούν τις πιο σημαντικές χαρακτηριστικές πληροφορίες των εικόνων, ονομάζονται "eigenfaces" και μπορούν να χρησιμοποιηθούν για αναγνώριση προσώπων.

Οι De et al. αρχικά εντόπισαν το πρόσωπο στην εικόνα με την μέθοδο HSV (Hue - Saturation - Value) και την απομόνωσαν, τμηματοποιώντας την εικόνα σε δύο μέρη: το μέρος του προσώπου και το φόντο. Έπειτα, εφάρμοσαν PCA για να πάρουν τα eigenfaces. Για την εκπαίδευση του μοντέλου ταξινόμησης των εικόνων του test set χρησιμοποίησαν την ευκλείδια απόσταση από το μέσο των eigenfaces του σετ εκπαίδευσης. Οι ερευνητές δεν ανέφεραν την συνολική ορθότητα του μοντέλου τους, ωστόσο αναφέρουν το ποσοστό αναγνώρισης του κάθε συναισθήματος ξεχωριστά. Το χαρούμενο συναίσθημα ήταν αυτό που απέφερε την μεγαλύτερη ορθότητα (93.1 %), ενώ το συναίσθημα του φόβου επέφερε την μικρότερη (77.7 %), σε ένα σετ εικόνων πιθανώς μη δημιοσιευμένο.

Οι Lajevardi και Lech [78] χρησιμοποίησαν λογαριθμικά φίλτρα Gabor για να προβλέψουν έξι συναισθήματα (θυμό, αηδία, φόβο, χαρά, θλίψη και έκπληξη) πάνω στις εικόνες του σετ CK [69], οι οποίες ουσιαστικά είναι καρέ από βίντεο. Καταρχάς, εντόπισαν το πρόσωπο σε κάθε εικόνα με την βοήθεια του γνωστού αλγόριθμου Viola-Jones [43]. Μετέπειτα, βρήκαν το καρέ του βίντεο κατά το οποίο το πρόσωπο είναι το πιο εκφραστικό, υπολογίζοντας την αμοιβαία πληροφορία (Mutual Information - MI) μεταξύ του αρχικού καρέ και του καρέ ενδιαφέροντος. Η αμοιβαία πληροφορία δύο διακριτών τυχαίων μεταβλητών X και Y υπολογίζεται ως εξής:

$$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2 \left(\frac{p(x,y)}{p(x) \cdot p(y)} \right)$$

Έπειτα, εφάρμοσαν λογαριθμικά φίλτρα Gabor και επέλεξαν το καλύτερο, και μετά επέλεξαν το καλύτερο σετ χαρακτηριστικών με βάση τον αλγόριθμο MIFS (Mutual Information based Feature Selection) ο οποίος βασίζεται στο κριτήριο της αμοιβαίας πληροφορίας. Στο τέλος, εκπαίδευσαν έναν ταξινομητή Naïve Bayes. Το σύστημα αυτό επέφερε 79.5 % ορθότητα στο σετ εικόνων CK [69].

Οι Alreshidi και Ullah [79] χρησιμοποίησαν μια τεχνική με όνομα "Neighbourhood Difference Features" (NDF) η οποία αξιοποιεί την σχέση γειτονικών περιοχών της εικόνας. Όπως και οι περισσότερες άλλες μέθοδοι, αρχικά ανιχνεύθηκε το πρόσωπο στην εικόνα χρησιμοποιώντας τον αλγόριθμο Viola-Jones [43], και στην συνέχεια εξήγαγαν τα χαρακτηριστικά NDF, τα οποία διατύπωσαν διαφορετικά μοτίβα με βάση τις σχέσεις μεταξύ γειτονικών περιοχών της εικόνας. Μετέπειτα, ταξινόμησαν τις εικόνες χρησιμοποιώντας Τυχαία Δέντρα (Random Forest Classifier) και προέβλεψαν

τα επτά συναισθήματα. Για να αξιολογήσουν το εν λόγω μοντέλο, χρησιμοποίησαν εκτός από την ορθότητα, την ανάκληση και την ακρίβεια. Συνολικά, το μοντέλο αυτό απέφερε 57.7 % ορθότητα στο σετ εικόνων SFEW [61] και 59 % ορθότητα στο RAF-DB [80].

Κεφάλαιο 4

Τεχνικές Λεπτομέρειες

4.1 Γλώσσα προγραμματισμού Python

Για όλες τις προγραμματιστικές εργασίες που αφορούν την παρούσα διπλωματική χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python και συγκεκριμένα η έκδοση 3.9.16. Η Python [81] [82] είναι μια γλώσσα προγραματισμού υψηλού επιπέδου η οποία χρησιμοποιείται ευρέως λόγω της απλότητας της σύνταξής της, καθώς και λόγω της πληθώρας βιβλιοθηκών (libraries) που είναι διαθέσιμες προς εγκατάσταση οι οποίες προσφέρουν αμέτρητες επιπρόσθετες λειτουργίες.

Η Python [81] [82] δημιουργήθηκε από τον Ολλανδό Γκίντο βαν Ρόσσουμ το 1989 (Guido van Rossum). Το όνομα της οφείλεται στην αγάπη του βαν Ρόσσουμ προς τους Βρετανούς κωμικούς Monty Python. Είναι μια διερμηνευόμενη γλώσσα προγραμματισμού (interpreted), και δεν απαιτεί προηγουμένως μεταγλώτισση (compilation), δηλαδή μετρατροπή του πηγαίου κώδικα σε κώδικα μηχανής.

Η Python χρησιμοποιείται σε ποικίλους τομείς. Πρωτίστως χρησιμοποιείται στον τομέα της τεχνητής νοημοσύνης και ιδιαίτερα στην μηχανική μάθηση, όπου μπορεί κανείς να αναπτύξει και να χρησιμοποιήσει μοντέλα μηχανικής μάθησης. Έπειτα, χρησιμοποιείται εκτενώς για την εκτέλεση διάφορων υπολογισμών, κυρίως επιστημονικών, όπως η ανάλυση δεδομένων, η στατιστική, ο απειροστικός λογισμός και η γραμμική άλγεβρα. Ακόμη, αξιοποιείται για την δημιουργία ιστοσελίδων και για την δημιουργία εφαρμογών.

Πιο συγκεκριμένα, εγκαταστάθηκε σε υπολογιστή με λειτουργικό σύστημα Windows μέσω του προγράμματος Anaconda [83], το οποίο έχει την δυνατότητα να δημιουργεί πολλαπλά ανεξάρτητα εικονικά περιβάλλοντα Python. Με την βοήθεια του Anaconda, κάθε περιβάλλον είναι απομονωμένο και προσαρμοσμένο στις ανάγκες για τις οποίες δημιουργήθηκε. Για παράδειγμα, ένα περιβάλλον που έχει ως στόχο την ανάπτυξη προγραμμάτων για ανάλυση εικόνων είναι πολύ πιθανόν να έχει διαφορετικές βιβλιοθήκες εγκατεστημένες από ένα περιβάλλον που έχει ως στόχο την

ανάλυση σημάτων.

Με την βοήθεια του Anaconda, δημιουργήθηκε ένα εικονικό περιβάλλον, το οποίο περιλαμβάνει εγκατεστημένες διάφορες βιβλιοθήκες που χρησιμοποιήθηκαν για την ανάπτυξη του κώδικα. Οι εν λόγω βιλιοθήκες θα αναφερθούν εκτενέστερα στο επόμενο υποκεφάλαιο. Για την ανάπτυξη και επεξεργασία του κώδικα χρησιμοποιήθηκε το Ολοκληρωμένο Περιβάλλον Ανάπτυξης (Integrated Development Environment - IDE) Visual Studio Code της Microsoft [84].

Επιπροσθέτως, προγράμματα τα οποία απαιτούσαν πολλή μνήμη και υπολογιστική δύναμη, όπως η εκπαίδευση τεχνητών νευρωνικών δικτύων, έτρεξαν στην πλατφόρμα Google Colab [85], η οποία παρέχει δωρεάν πρόσβαση σε μονάδες επεξεργασίας γραφικών (GPU) και μονάδες επεξεργασίας Tensor (TPU), οι οποίες προσφέρουν επιτάχυνση στην εκτέλεση αλγόριθμων μηχανικής μάθησης. Το όνομα του Google Colab είναι σύντμηση της λέξης Colaboratory διότι προσφέρει online πρόσβαση σε πηγαίο κώδικα επιτρέποντας την ταυτόχρονη επεξεργασία του από πολλαπλούς χρήστες. Τέλος, προσφέρει δωρεάν αποθήκευση αρχείων, ή ακόμα και εύκολη πρόσβαση στα αρχεία του χρήστη που βρίσκονται αποθηλευμένα στο Google Drive.

4.2 Τύποι αρχείων

Ο συνήθης τύπος αρχείων Python είναι τα αρχεία με κατάληξη ".py". Τα αρχεία αυτά περιέχουν μόνο πηγαίο κώδικα Python. Χρησιμοποιούνται για την ανάπτυξη και εκτέλεση προγραμμάτων Python. Ωστόσο, υπάρχει ένας επιπλέον τύπος αρχείων με κατάληξη ".ipynb". Τα αρχεία τύπου ".ipynb" είναι αρχεία Jupyter Notebook και είναι διαδραστικά αρχεία που περιέχουν κελιά με κώδικα Python αλλά και εικόνες, κελιά με κείμενο, τίτλους και άλλα στοιχεία. Η ειδοποιός διαφορά τους είναι πως το δεύτερο δίνει την επιλογή της εκτέλεσης μόνο των επιθυμητών κελιών, καθώς και την άμεση προβολή των αποτελεσμάτων σε πραγματικό χρόνο. Όλος ο πηγαίος κώδικας που αναπτύχθηκε βρίσκεται σε αρχεία τύπου ".ipynb".

4.3 Βιβλιοθήκες

Παρακάτω αναγράφονται οι βιβλιοθήκες που χρησιμοποιήθηκαν για την επεξεργασία εικόνων, για τον χειρισμό των δεδομένων, για την ανάπτυξη των μοντέλων μηχανικής μάθησης και για την αξιολόγησή τους.

- **Numpy [86]**: Παρέχει λειτουργίες για την επιστημονική και αριθμητική επεξεργασία δεδομένων, καθώς και λειτουργίες γραμμικής άλγεβρας για πράξεις μεταξύ πινάκων. Δεδομένου ότι οι εικόνες διαβάζονται από τον υπολογιστή ως δισδιάστατοι ή τρισδιάστατοι πίνακες αριθμών, όπου κάθε αριθμός αποτελεί την ένταση του κάθε εικονοστοιχείου, η βιβλιοθήκη Numpy αποτελεί χρήσιμο εργαλείο για τον χειρισμό των εικόνων ως αριθμητικοί πίνακες.

- **Pandas [87]**: Προσφέρει εργαλεία εισαγωγής, ανάλυσης και επεξεργασίας δεδομένων. Προσφέρει την ευρέως χρησιμοποιούμενη δομή δεδομένων DataFrame.

- **Tensoflow [88]**: Παρέχει ένα πλαίσιο για την δημιουργία, εκπαίδευση και αξιολόγηση τεχνητών νευρωνικών δικτύων. Επίσης, παρέχει συναρτήσεις που βοηθούν στην ταχεία εκτέλεση των αλγόριθμων στον επεξεργαστή της κάρτας γραφικών (GPU).

- **Keras [89]**: Το Keras αρχικά αναπτύχθηκε για την κατασκευή τεχνητών νευρωνικών δικτύων. Αργότερα, όμως, αφομοιώθηκε στην βιβλιοθήκη Tensorflow και χρησιμοποιείται ως διεπαφή (interface) για την δημιουργία και εκπαίδευση τους.

- **OpenCV [90]**: Προσφέρει εργαλεία για την επεξεργασία εικόνων αλλά και πολλαπλές δυνατότητες που αφορούν την υπολογιστική όραση. Το OpenCV χρησιμοποιήθηκε στην παρούσα εργασία για τον εντοπισμό προσώπων στις εικόνες.

- **Scikit-learn [91]**: Παρέχει εργαλεία για την ανάλυση και προ-επεξεργασία των δεδομένων για να εισαχθούν αργότερα σε μοντέλα μηχανικής μάθησης. Επίσης, προσφέρει τα εργαλεία για την δημιουργία, εκπαίδευση και αξιολόγηση των εν λόγω μοντέλων.

- **Matplotlib [92]**: Πρόκειται για μια βιβλιοθήκη οπτικοποίησης η οποία παρέχει όλες τις απαραίτητες λειτουργίες για την παραγωγή γραφημάτων κάθε τύπου. Επίσης, χρησιμοποιείται για την οπτικοποίηση των εικόνων, επεξεργασμένων και μη, καθώς και για την εμφάνιση πολλαπλών υπο-γραφημάτων ή εικόνων σε ένα ενιαίο γράφημα.

Κεφάλαιο 5

Datasets εικόνων

5.1 KDEF

Το σετ εικόνων Karolinska Directed Emotional Faces (KDEF) [64] είναι μια συλλογή 2940 έγχρωμων εικόνων που συμπεριλαμβάνει επτά συναισθήματα (χαρά, λύπη, θυμό, αηδία, φόβο, έκπληξη και ουδέτερη έκφραση). Το σετ αυτό σημιουργήθηκε το 1998 από το Ινστιτούτο Καρολίνσκα στην Σουηδία και περιλαμβάνει εικόνες από 70 άτομα, εκ των οποίων 35 είναι γυναίκες και 35 είναι άντρες, με διαφορετικές γωνίες λήψης.

Κάθε πρόσωπο απεικονίζεται με ένα συναίσθημα μόνο μία φορά, με σκοπό να επιβεβαιωθεί ότι το συναίσθημα είναι αυθεντικό. Το KDEF είναι πολύτιμος πόρος για την έρευνα πάνω σε εικόνες, ειδικά στον τομέα της αναγνώρισης συναισθημάτων και της ανάλυσης εκφράσεων του προσώπου. Στην εικόνα 5.1 απεικονίζεται ένα τυχαία επιλεγμένο παράδειγμα εικόνας από κάθε κλάση, δηλαδή από κάθε συναίσθημα. Τέλος, όλες οι εικόνες είναι διαθέσιμες μέσω αρχείου JPG.

5.2 JAFFE

Το σετ εικόνων Japanese Female Facial Expression (JAFFE) [45] [44] αποτελείται από 213 ασπρόμαυρες εικόνες που εκφράζουν επίσης επτά συναισθήματα (χαρά, λύπη, θυμό, αηδία, φόβο, έκπληξη και ουδέτερη έκφραση). Περιλαμβάνει 213 ει-

Σχήμα 5.1: Παραδείγματα εικόνων από το σετ KDEF

Σχήμα 5.2: Παραδείγματα εικόνων από το σετ JAFFE

κόνες που καταγράφουν διάφορες εκφράσεις από ένα σύνολο 10 γυναικών. Το σύνολο εικόνων JAFFE αναπτύχθηκε για ερευνητικούς σκοπούς και είναι προσβάσιμο για λήψη μέσω του διαδικτύου. Στην εικόνα 5.2 απεικονίζεται ένα τυχαία επιλεγμένο παράδειγμα εικόνας από κάθε κλάση, δηλαδή από κάθε συναίσθημα. Όλες οι εικόνες είναι διαθέσιμες μέσω αρχείου TIFF (Tagged Image File Format).

5.3 Διερευνητική ανάλυση εικόνων

Σε αυτό το υποκεφάλαιο θα παρουσιασθούν κάποια βασικά διερευνητικά βήματα τα οποία έγιναν και στα δυο σετ εικόνων, με σκοπό να δούμε πώς μοιάζει το κάθε σετ συνολικά αλλά και το κάθε συναίσθημα ξεχωριστά.

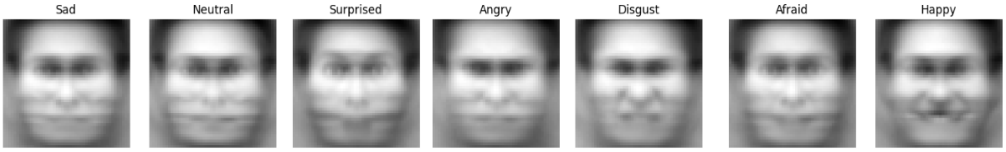
Αρχικά, για κάθε σετ εικόνων υπολογίσθηκε η "μέση" εικόνα κάθε συναισθήματος. Πιο συγκεκριμένα, βρέθηκε ο μέσος όρος των εικονοστοιχείων όλων των εικόνων για κάθε ξεχωριστό συναίσθημα. Ο μέσος όρος των εικόνων λειτουργεί ως μια συνθετική απεικόνιση των βασικών χαρακτηριστικών προσώπου που συσχετίζονται με ένα συγκεκριμένο συναίσθημα. Το "μέσο πρόσωπο" καταγράφει τα οπτικά στοιχεία και τα χαρακτηριστικά που συνήθως συσχετίζονται με ένα συγκεκριμένο συναίσθημα και ουσιαστικά παρέχει μια συνοπτική αναπαράσταση των χαρακτηριστικών προσώπου που σχετίζονται με αυτό.

Για το KDEF μπορούμε να παρατηρήσουμε πως στην μέση εικόνα των εικόνων μπροστινής λήψης τα συναισθήματα είναι αρκετά διακριτά, ακόμα και αν η μέση εικόνα κάθε συναισθήματος είναι πιο θολή. Το ίδιο ισχύει και για το JAFFE, στο οποίο διακρίνεται εύκολα και το φύλο των συμμετεχόντων, αφού το σετ αυτό αποτελείται αποκλειστικά από γυναίκες.

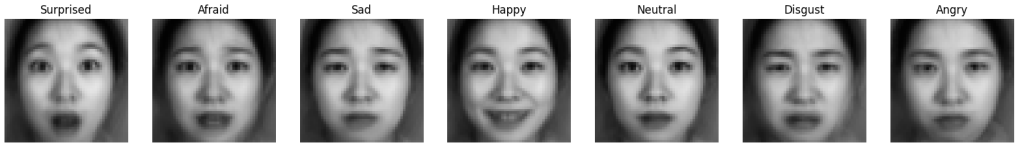
Η διαδικασία βοηθά στη μείωση του θορύβου, παρέχοντας μια πιο καθαρή απεικόνιση των βασικών χαρακτηριστικών του προσώπου που σχετίζονται με το κάθε συναίσθημα. Τα αποτελέσματα οπτικοποιούνται στην εικόνα 5.3.

Έπειτα, παρατηρήσαμε την συχνότητα του εκάστοτε συναισθήματος στο κάθε σετ εικόνων. Οι συχνότητες απεικονίζονται σε ραβδογράμματα στις εικόνες 5.4 και 5.5. Το βήμα αυτό ήταν σημαντικό για να παρατηρήσουμε αν υπήρχαν ανισορροπίες στον αριθμό των συναισθημάτων, δηλαδή αν είχαμε μεγάλες διαφορές μεταξύ των ποοσοτήτων των εικόνων που ανήκουν σε κάθε έκφραση.

(α') KDEF (μόνο μπροστινές εικόνες)

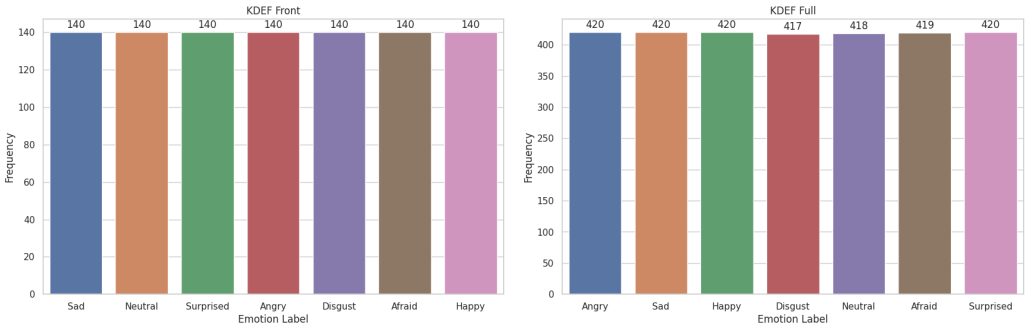


(β') KDEF (μαζί με εικόνες υπό γωνία)



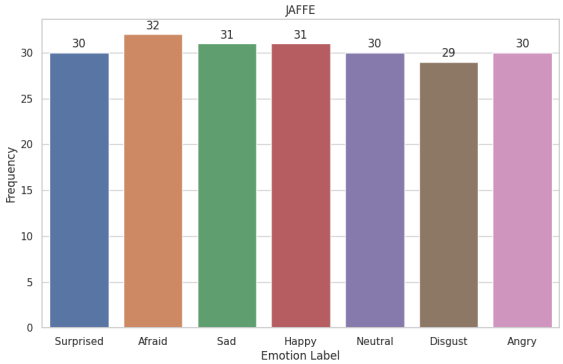
(γ') JAFFE

Σχήμα 5.3: Οπτικοποίηση μέσου προσώπου για κάθε συναίσθημα



(α') Χωρίς τις εικόνες υπό γωνία (β') Με τις εικόνες υπό γωνία

Σχήμα 5.4: Ραβδόγραμμα συχνοτήτων συναισθήματος για το σετ εικόνων KDEF



Σχήμα 5.5: Ραβδόγραμμα συχνοτήτων συναισθήματος για το σετ εικόνων JAFFE

5.4 Ομοιότητα συναισθημάτων

Στο πλαίσιο της εξερεύνησης και αρχικής ανάλυσης των εικόνων υπολογίσθηκε και η ομοιότητα μεταξύ των συναισθημάτων για κάθε σετ εικόνων ξεχωριστά. Η ομοιότητα των εικόνων είναι μια ευρέως χρησιμοποιούμενη μέθοδος η οποία βοηθά στην εύρεση παρόμοιων ειικόνων με σκοπό την εξαγωγή των κοινών τους χαρακτηριστικών [93]. Υπάρχουν διάφορες συναρτήσεις που μετρούν την ομοιότητα (similarity functions) [94] [95] [96], ωστόσο σε αυτή την περίπτωση χρησιμοποιήθηκε η ευκλείδια απόσταση. Καθώς θα ήταν πιο πολύπλοκο να υπολογισθεί η ομοιότητα μεταξύ όλων των εικόνων, χρησιμοποιήθηκε η μέση εικόνα κάθε συναισθήματος.

Αρχικά, υπολογίσθηκε η μέση εικόνα, όπως περιγράφηκε στο προηγούμενο κεφάλαιο. Έπειτα, μετρήθηκε η ευκλείδια απόσταση των εικονοστοιχείων μεταξύ κάθε ζεύγους συναισθημάτων [97]. Η ευκλείδια απόσταση μετρά την "απόσταση" μεταξύ δύο σημείων σε έναν πολυδιάστατο χώρο και μπορεί να χρησιμοποιηθεί για να εκτιμήσουμε πόσο διαφορετικά είναι δύο διανύσματα. Τέλος, οι τιμές των αποστάσεων κανονικοποιήθηκαν στο [0, 1], χρησιμοποιώντας τον παρακάτω τύπο:

$$similarity\ matrix_{normalized} = 1 - \frac{similarity\ matrix}{max(similarity\ matrix)}$$

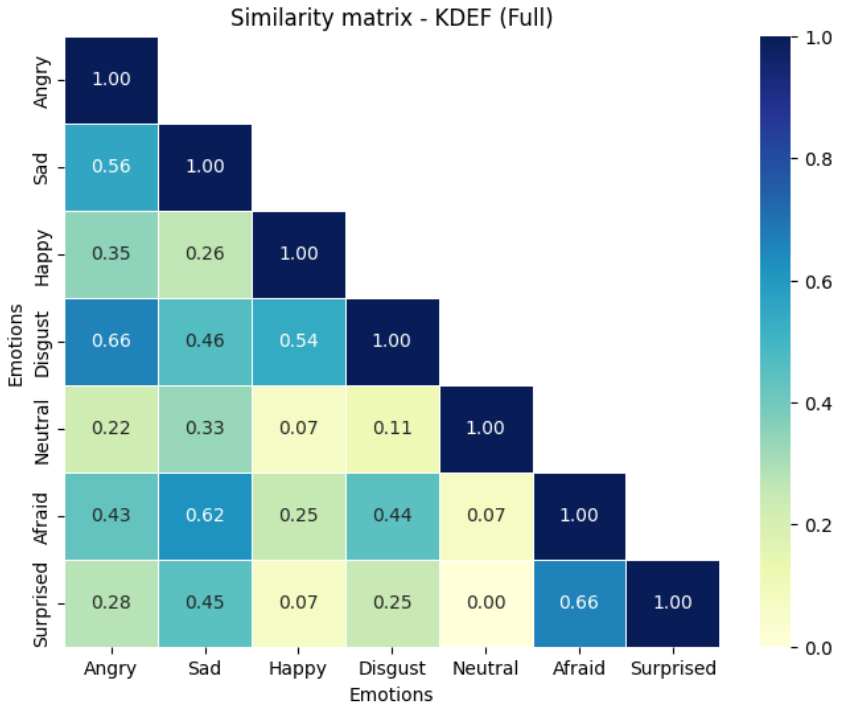
όπου $max(similarity\ matrix)$ η μέγιστη τιμή του πίνακα.

Το αποτέλεσμα ήταν ένας πίνακας ομοιότητας, ο οποίος περιέχει τιμές που αντανακλούσαν το βαθμό ομοιότητας μεταξύ όλων των συναισθημάτων. Αυτός ο πίνακας με βάση την ευκλείδια απόσταση, αποτελεί ένα εργαλείο που βοηθά στην κατανόηση των σχέσεων μεταξύ των συναισθημάτων αναδεικνύει πιθανές ομάδες συναισθημάτων με παρόμοια χαρακτηριστικά.

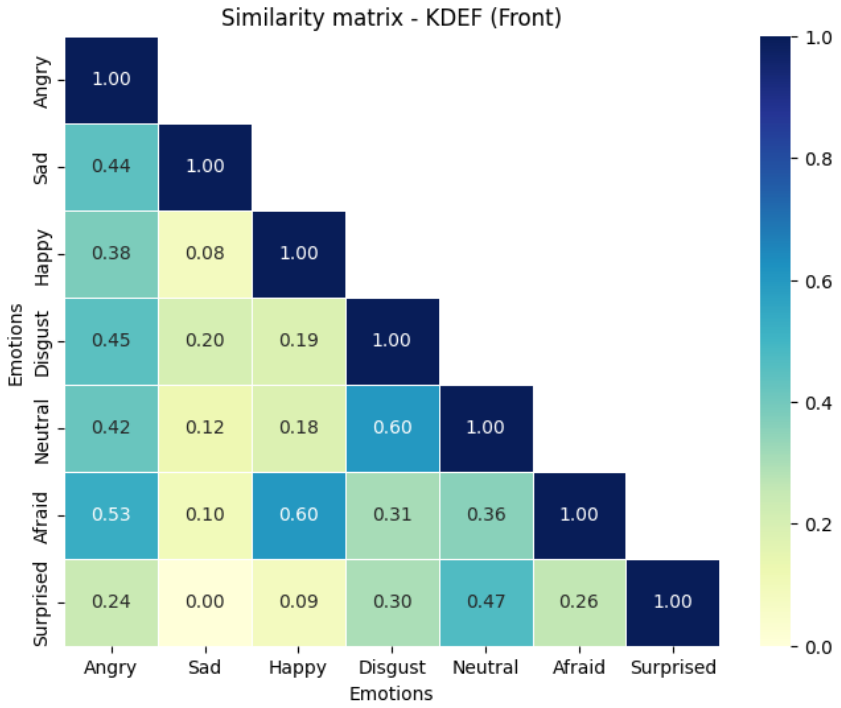
Οι τελικοί πίνακες ομοιότητας παρουσιάζονται ως heatmap στις εικόνες 5.6, 5.7 και 5.8.

Όπως παρατηρείται και στις εικόνες, υπάρχουν ομοιότητες μεταξύ των συναισθημάτων, με μέγιστο βαθμό ομοιότητας 0.66. Πιο συγκεκριμένα, στις μπροστινές εικόνες του KDEF παρατηρείται μεγάλη ομοιότητα του θυμού με σχεδόν όλα τα συναισθήματα, ιδιαίτερα με τα αρνητικά, αλλά κυρίως με τον φόβο. Επίσης, υπάρχει ομοιότητα μεταξύ της χαράς και του φόβου, η οποία θα μπορούσε να δικαιολογηθεί, καθώς και τα δύο συναισθήματα εκφράζονται συχνά με ανοιχτό στόμα. Επιπροσθέτως, υπάρχει υψηλή ομοιότητα του ουδέτερου συναισθήματος με το συναίσθημα της αηδίας.

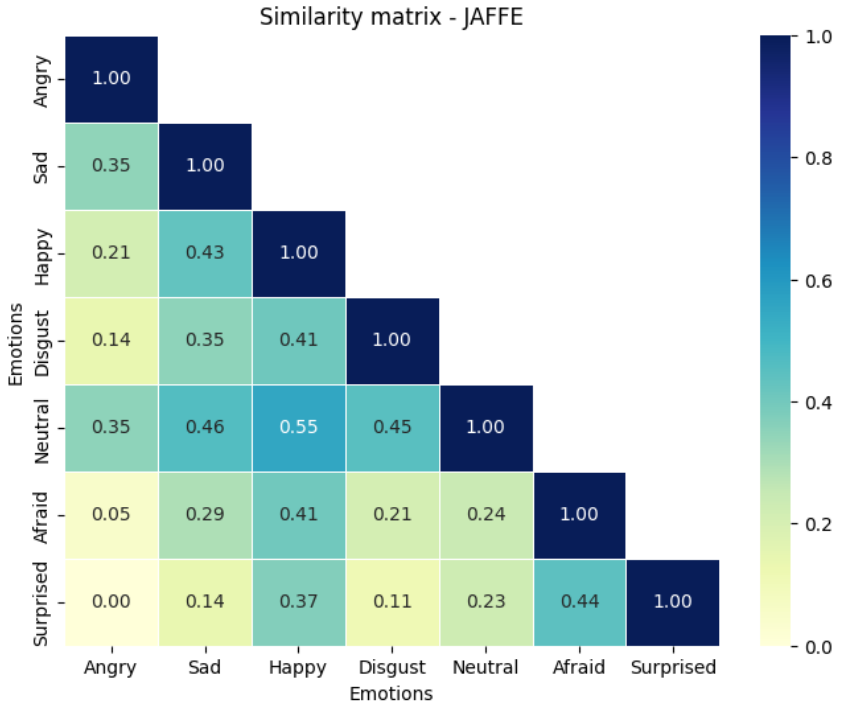
Στις εικόνες του KDEF υπό γωνία λήψης παρατηρείται μεγάλη ομοιότητα μεταξύ του θυμού και της λύπης καθώς και της αηδίας, του φόβου και της έκπληξης, αλλά και της λύπης με τον φόβο. Το πιο αξιοσημείωτο είναι η ομοιότητα της αηδίας με το χαρούμενο συναίσθημα. Στις εικόνες του JAFFE, παρατηρείται ομοιότητα μεταξύ της χαράς και του ουδέτερου συναισθήματος. Η ομοιότητα μεταξύ των συναισθημάτων σε



Σχήμα 5.6: Πίνακας ομοιότητας συναισθημάτων για το KDEF (Full)



Σχήμα 5.7: Πίνακας ομοιότητας συναισθημάτων για το KDEF (Front)



Σχήμα 5.8: Πίνακας ομοιότητας συναισθημάτων για το JAFFE

κάθε εικόνα θα μπορούσε να εξηγήσει πιθανές λάθος προβλέψεις του μοντέλου.

Κεφάλαιο 6

Μεθοδολογία

Στο συγκεκριμένο κεφάλαιο θα εξετασθούν οι αλγόριθμοι επιβλεπόμενης μάθησης και πιο συγκεκριμένα τα τεχνητά νευρωνικά δίκτυα τα οποία χρησιμοποιήθηκαν για την ταξινόμηση των συναισθημάτων, καθώς και οι μέθοδοι προεπεξεργασίας των εικόνων. Θα αναφερθούν οι αρχιτεκτονικές τεχνητών νευρωνικών δικτύων που αναπτύχθηκαν για την ταξινόμηση συναισθημάτων από εικόνες προσώπων και θα εξετασθεί πώς αυτές οι αρχιτεκτονικές διαμορφώνονται για να αντιμετωπίσουν την πρόκληση της αναγνώρισης και ταξινόμησης των συναισθημάτων. Θα αναλυθεί η δομή, οι συνιστώσες και οι συναρτήσεις ενεργοποίησης που χρησιμοποιούνται σε κάθε αρχιτεκτονική, εξηγώντας πώς αυτά τα στοιχεία συμβάλλουν στην εξαγωγή χαρακτηριστικών από τις εικόνες.

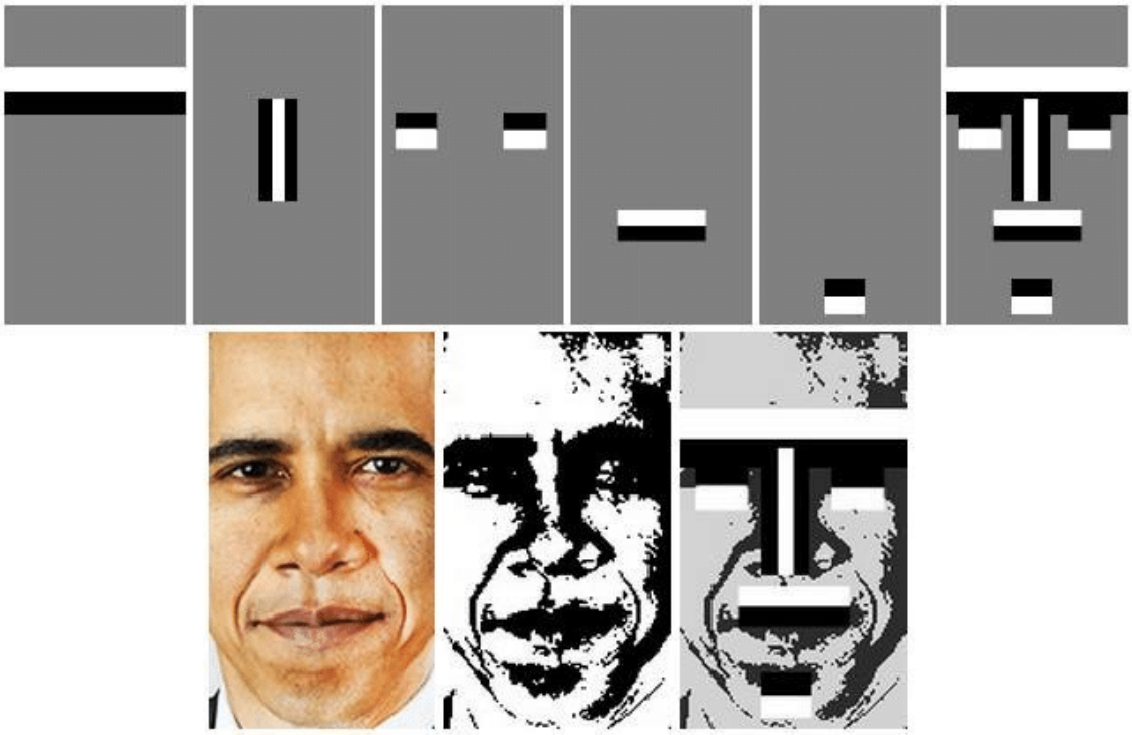
6.1 Προεπεξεργασία εικόνων

Όλες οι εικόνες πέρασαν από κάποια βήματα προεπεξεργασίας πριν μπουν ως είσοδος στα εκάστοτε νευρωνικά δίκτυα. Αρχικά, σε όλες τις εικόνες εφαρμόσθηκε ένας αλγόριθμος εντοπισμού προσώπου, με σκοπό να απομονωθεί το πρόσωπο και να καταλάβει το μεγαλύτερο μέρος της εικόνας, χωρίς να υπάρχει περιττός "θόρυβος" από άλλα αντικείμενα ή από το φόντο στο βάθος. Οι αλγόριθμοι εύρεσης προσώπου συνήθως επιστρέφουν τις συντεταγμένες ενός τετραγώνου στον χώρο της εικόνας, εντός του οποίου εμπεριέχεται το πρόσωπο.

6.1.1 Αλγόριθμος Viola-Jones

Σε όλες τις μπροστινές εικόνες προσώπων και των δύο datasets εφαρμόστηκε ο αλγόριθμος εύρεσης προσώπου Viola Jones [43]. Πρόκειται για έναν ευρέως διαδεδομένο αλγόριθμο ο οποίος δημιουργήθηκε το 2001 από τους ερευνητές Paul Viola και Michael Jones, από τους οποίους και πήρε το όνομά του. Ο αλγόριθμος βασίζεται σε έναν συνδυασμό τεχνικών μηχανικής μάθησης και επεξεργασίας εικόνας.

Ο αλγόριθμος Viola-Jones λειτουργεί χρησιμοποιώντας μια προσέγγιση που βα-



Σχήμα 6.1: Χαρακτηριστικά Haar (Haar-like features)

σίζεται σε χαρακτηριστικά Haar (Haar-like features). Τα χαρακτηριστικά Haar είναι ορθογώνια μοτίβα που καταγράφουν τις τοπικές παραλλαγές εικόνας, τα οποία είναι απλά και αποδοτικά στον υπολογισμό (βλ. εικόνα 6.1). Ο αλγόριθμος χρησιμοποιεί ένα σύνολο τέτοιων χαρακτηριστικών για να ταξινομήσει τις περιοχές εικόνας είτε ως "πρόσωπο" ή "μη πρόσωπο" με βάση τα μοτίβα έντασης τους.

Ο αλγόριθμος Viola-Jones αποτελείται από πολλά βασικά βήματα. Πρώτον, χρησιμοποιεί μια τεχνική που ονομάζεται integral images για να επιταχύνει τον υπολογισμό χαρακτηριστικών Haar. Οι integral images επιτρέπουν να υπολογιστούν αποτελεσματικά τα αθροίσματα των εντάσεων των εικονοστοιχείων σε οποιαδήποτε ορθογώνια περιοχή μιας εικόνας. Αυτό επιτρέπει την ταχεία αξιολόγηση χαρακτηριστικών.

Στη συνέχεια, ο αλγόριθμος χρησιμοποιεί έναν μοντέλο ταξινόμησης που βασίζεται στον αλγόριθμο AdaBoost. Ο AdaBoost είναι ένας αλγόριθμος μηχανικής μάθησης που συνδυάζει πολλούς αδύναμους ταξινομητές σε έναν ισχυρό ταξινομητή. Στον αλγόριθμο Viola-Jones, κάθε ασθενές μοντέλο ταξινόμησης αντιστοιχεί σε ένα χαρακτηριστικό Haar. Ο αλγόριθμος AdaBoost χρησιμοποιείται για να επιλέξει τα πιο χρήσιμα χαρακτηριστικά και για να εκχωρήσει τα κατάλληλα βάρη σε αυτά.

Κατά τη φάση ανίχνευσης, ο αλγόριθμος σαρώνει την εικόνα με ένα κινούμενο παράθυρο διαφόρων μεγεθών, εφαρμόζοντας τα χαρακτηριστικά Haar σε κάθε παράθυρο. Σε κάθε βήμα, ο αλγόριθμος αξιολογεί τα χαρακτηριστικά Haar χρησιμοποιώντας το ισχυρό μοντέλο ταξινόμησης που εκπαιδεύεται από το AdaBoost. Εάν

Σχήμα 6.2: Παράδειγμα εικόνων KDEF σε λήψη υπό γωνία

μια περιοχή ταξινομηθεί ως πρόσωπο, επαληθεύεται περαιτέρω χρησιμοποιώντας και άλλα μοντέλα ταξινόμησης για τη μείωση των False Positives.

Ο αλγόριθμος Viola-Jones είναι γνωστός για τα υψηλά ποσοστά επιτυχίας ανίχνευσης και τις χαμηλές υπολογιστικές του απαιτήσεις. Έχει χρησιμοποιηθεί ευρέως σε εφαρμογές όπως η ανίχνευση προσώπου σε ψηφιακές κάμερες, συστήματα παρακολούθησης βίντεο και ανάλυση εκφράσεων προσώπου σε πραγματικό χρόνο. Ωστόσο, μπορεί να έχει περιορισμούς όταν πρόκειται για την ανίχνευση προσώπων σε κάποιες γωνίες, τις συνθήκες φωτισμού ή με εμπόδια. Ο αλγόριθμος Viola-Jones έχει συνεισφέρει σημαντικά στον τομέα της όρασης υπολογιστών και έχει ανοίξει το δρόμο για περαιτέρω πρόοδο στον εντοπισμό και την αναγνώριση προσώπου.

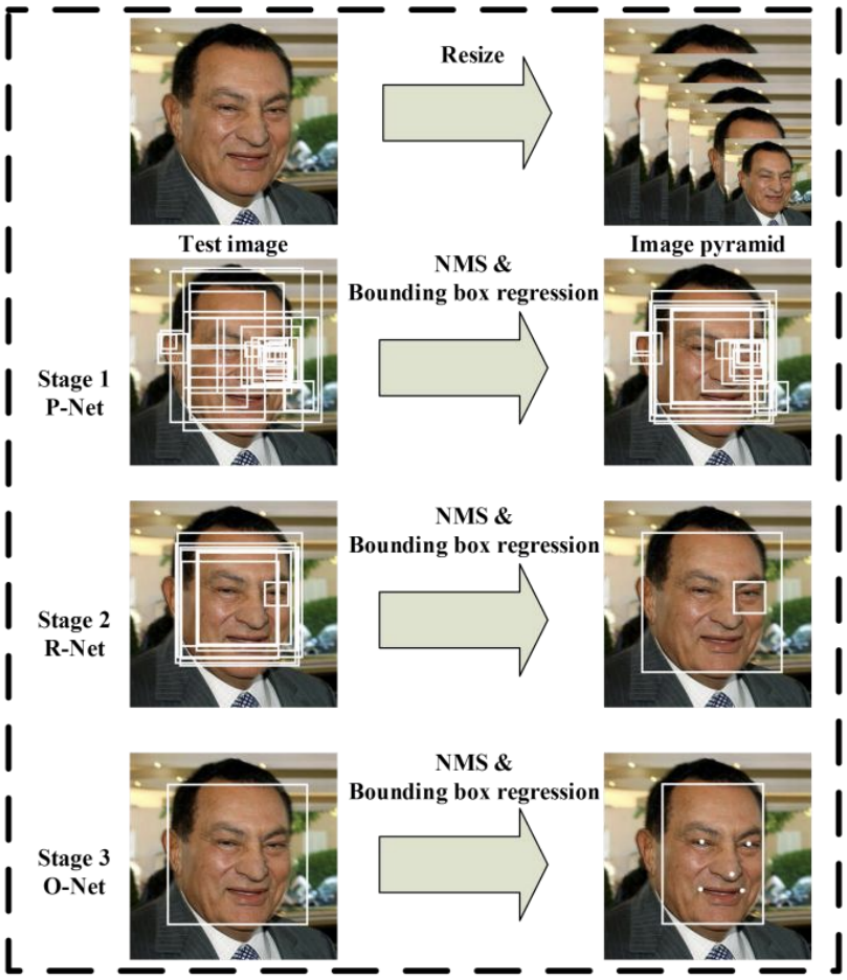
6.1.2 Αλγόριθμος MTCNN

Στις εικόνες του dataset KDEF που είχαν ληφθεί υπό γωνία (βλ. εικόνα 6.2), δηλαδή που δεν είχαν ολόκληρη την εικόνα του προσώπου από μπροστά, εφαρμόσθηκε ο αλγόριθμος MTCNN (Multi-Task Cascaded Convolutional Neural Networks) [4]. Ο MTCNN είναι ένας ευρέως διαδεδομένος αλγόριθμος ανίχνευσης προσώπων σε εικόνες ο οποίος στοχεύει στο να βρει την θέση των χαρακτηριστικών του προσώπου σε μια εικόνα. Είναι γνωστός για την ανθεκτικότητά του στις αλλαγές της θέσης και της κλίσης του προσώπου και έχει ευρεία εφαρμογή σε περιπτώσεις όπως η εύρεση συναισθημάτων και η ανίχνευση χαρακτηριστικών προσώπου.

Αποτελείται από τρία διαδοχικά στάδια τα οποία εκτελούνται για να βελτιστοποιηθεί το τελικό αποτέλεσμα. Το πρώτο στάδιο αποτελείται από ένα συνελικτικό νευρωτικό δίκτυο (CNN) το οποίο ονομάζεται P-Net (Proposal Network) που βρίσκει πιθανές περιοχές προσώπου σε μια εικόνα. Αυτό το στάδιο παράγει ένα σετ πιθανών τετραγώνων ή "κουτιών" οριοθέτησης (bounding boxes) τα οποία μπορεί να περιέχουν πρόσωπο.

Στο δεύτερο στάδιο, τα τετράγωνα υπόκεινται σε περαιτέρω έλεγχο, με τη χρήση ενός άλλου (CNN) το οποίο ονομάζεται R-Net (Refine Network). Αυτό το στάδιο βελτιώνει τα κουτιά οριοθέτησης, προσαρμόζοντας τις συντεταγμένες τους, για να ευθυγραμμιστούν καλύτερα με τα πρόσωπα. Επιπροσθέτως, παράγει τα σημεία-ορόσημο του κάθε προσώπου, που υποδεικνύουν την θέση των διαφόρων χαρακτηριστικών, όπως το στόμα, τα μάτια και η μύτη.

Στο τελευταίο στάδιο, χρησιμοποιείται ένα άλλο (CNN), το λεγόμενο O-Net (Output



Σχήμα 6.3: Γραφική απεικόνιση του αλγορίθμου MTCNN, όπως παρουσιάσθηκε στο αντίστοιχο ερευνητικό άρθρο [4]

Network), το οποίο ταξινομεί τα κουτιά οριοθέτησης σαν πρόσωπα ή μη-πρόσωπα. Κατά την διάρκεια της διαδικασίας αυτής απορρίπτει τα ψευδώς ανιχνευμένα πρόσωπα και δίνει ως τελικό αποτέλεσμα αυτά που βρήκε, μαζί με τα σημεία-ορόσημο του κάθε προσώπου.

Οι δύο προαναφερθείσες μέθοδοι εντοπισμού προσώπου σε εικόνα κρίθηκαν χρήσιμοι κατά την προεπεξεργασία των δεδομένων και παρήγαγαν αξιόπιστα αποτελέσματα.

6.1.3 Άλλα βήματα προεπεξεργασίας

Αφού απομονώθηκε το πρόσωπο από κάθε εικόνα, οι εικόνες πέρασαν από μερικά ακόμη στάδια προεπεξεργασίας. Πριν από όλα, είναι σημαντικό να αναφερθεί πως το σετ KDEF περιείχε έξι εικόνες οι οποίες δεν είχαν πρόσωπα και ήταν τελείως μαύρες. Πρώτα αυτές αφαιρέθηκαν, και μετά οι υπόλοιπες μετατράπηκαν σε ασπρόμαυρες (grayscale) και μετατράπηκαν σε μέγεθος (48,48). Η προεπεξεργασία αυτή έγινε με σκοπό την επιτάχυνση της διαδικασίας εκπαίδευσης και ελέγχου. Έπειτα, οι τιμές

Σχήμα 6.4: Διαδικασία προεπεξεργασίας εικόνων

των εικονοστοιχείων της κάθε εικόνας διαιρέθηκαν με το 255, δηλαδή την μέγιστη τιμή που μπορεί να πάρει ένα pixel. Η διαδικασία αυτή ονομάζεται κανονικοποίηση και κλιμακώνει τις τιμές των εικονοστοιχείων από το αρχικό εύρος (από 0 έως 255) σε ένα νέο εύρος από 0 έως 1. Αυτό το εύρος είναι πιο κατάλληλο για πολλούς αλγόριθμους μηχανικής εκμάθησης, επειδή διασφαλίζει πως όλα τα χαρακτηριστικά (τιμές εικονοστοιχείων σε αυτήν την περίπτωση) έχουν την ίδια κλίμακα. Αλγόριθμοι, όπως τα νευρωνικά δίκτυα, συγκλίνουν πιο γρήγορα όταν τα χαρακτηριστικά εισόδου είναι σε παρόμοια κλίμακα. Επίσης, η κανονικοποίηση βοηθάει να αποφευχθεί το πρόβλημα των εξαφανιζόμενων παραγώγων (vanishing gradients) όταν χρησιμοποιείται σιγμοειδής συνάρτηση ενεργοποίησης.

6.2 Σύνολο εικόνων

Όπως αναπτύχθηκε και στο προηγούμενο υποκεφάλαιο, τα δύο σετ εικόνων που χρησιμοποιήθηκαν για την εκπαίδευση και επιβεβαίωση των μοντέλων ήταν τα Karolinska Directed Emotional Faces (KDEF) [64] και Japanese Female Facial Expression (JAFFE) [45] [44], τα οποία και προεπεξεργάσθηκαν σύμφωνα με τις μεθόδους που αναφέρθηκαν στο υποκεφάλαιο 6.1.

Για το σετ εικόνων KDEF, τα μοντέλα που δημιουργήθηκαν εφαρμόσθηκαν σε όλο το σετ αλλά και ξεχωριστά για τις εικόνες που ήταν μόνο τραβηγμένες από μπροστά, χωρίς να συμπεριληφθούν οι εικόνες προσώπων υπό γωνία. Για το σετ εικόνων JAFFE δεν έγινε κάποιος τέτοιος διαχωρισμός καθώς δεν περιέχει εικόνες υπό μεγάλη γωνία λήψης αλλά ούτε και το μικρό του μέγεθος το επιτρέπει.

6.3 Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου

Σε αυτό το υποκεφάλαιο περιγράφεται το βασικό μοντέλο επιβλεπόμενης μάθησης το οποίο χρησιμοποιήθηκε για την ταξινόμηση των διαφόρων συναισθημάτων. Πρόκειται για ένα τεχνητό νευρωνικό δίκτυο το οποίο χρησιμοποιεί συνελικτικά στρώματα. Επομένως, αποκαλείται συνελικτικό και όπως και σε προηγούμενα κεφάλαια, θα αναφέρεται εν συντομία ως CNN.

Το βασικό δίκτυο (βλ. εικόνα 6.5) αποτελείται συνολικά από δέκα στρώματα. Πιο συγκεκριμένα αποτελείται από:

1. Ένα **στρώμα εισόδου** που δέχεται εικόνες μεγέθους $(48, 48, 1)$, με το 1 να δείχνει το grayscale κανάλι καθώς οι εικόνες είναι grayscale.

2. Ένα δισδιάστατο **συνελικτικό στρώμα** με 32 φίλτρα διαστάσεων $(3, 3)$ και συνάρτηση ενεργοποίησης ReLU.

3. Ένα δισδιάστατο **στρώμα Max Pooling** με pool size $(2, 2)$ το οποίο μειώνει το μέγεθος του χάρτη χαρακτηριστικών (feature map) που βγήκε από το προηγούμενο στρώμα.

4. Ένα δισδιάστατο **συνελικτικό στρώμα** με 64 φίλτρα διαστάσεων $(3, 3)$ και συνάρτηση ενεργοποίησης ReLU.

5. Ένα ακόμη δισδιάστατο **στρώμα Max Pooling** με pool size $(2, 2)$

6. Ένα δισδιάστατο **συνελικτικό στρώμα** με 128 φίλτρα διαστάσεων $(3, 3)$ και συνάρτηση ενεργοποίησης ReLU.

7. Ένα ακόμη δισδιάστατο **στρώμα Max Pooling** με pool size $(2, 2)$

8. Ένα στρώμα **Flatten** το οποίο μεταμορφώνει την είσοδο του "ισοπεδώνοντας" την έξοδο από τα προηγούμενα συνελικτικά στρώματα σε ένα μονοδιάστατο διάνυσμα, προετοιμάζοντας το για τα πλήρως συνδεδεμένα στρώματα που ακολουθούν.

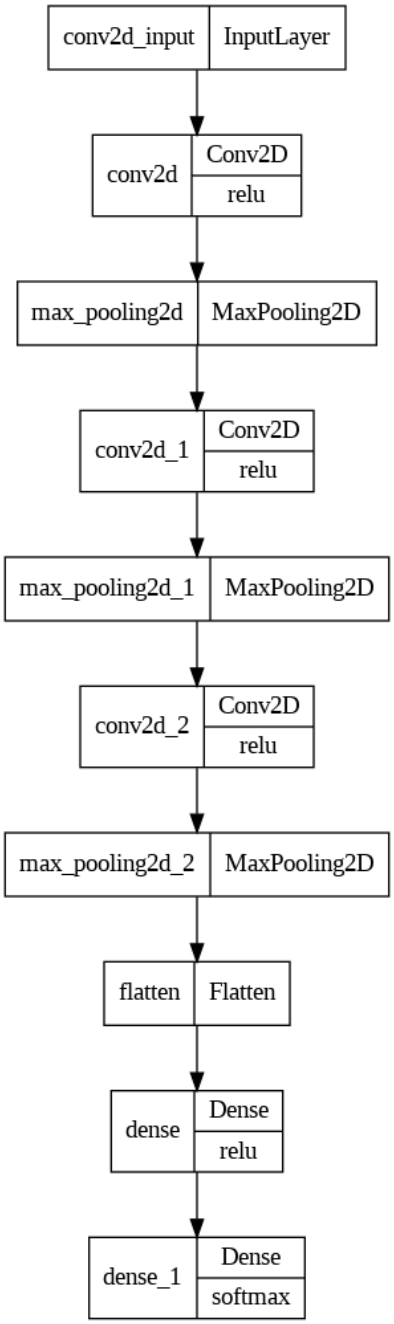
9. Ένα **πλήρως συνδεδεμένο στρώμα** με 128 νευρώνες και συνάρτηση ενεργοποίησης ReLU.

10. Ένα **πλήρως συνδεδεμένο στρώμα** με 7 νευρώνες (όσο και ο αριθμός των πιθανών εκφράσεων του προσώπου) και συνάρτηση ενεργοποίησης SoftMax, που είναι και το στρώμα εξόδου. Η συνάρτηση ενεργοποίησης SoftMax μετατρέπει την έξοδο σε κατανομή πιθανοτήτων που δείχνει την πιθανότητα του κάθε συναισθήματος.

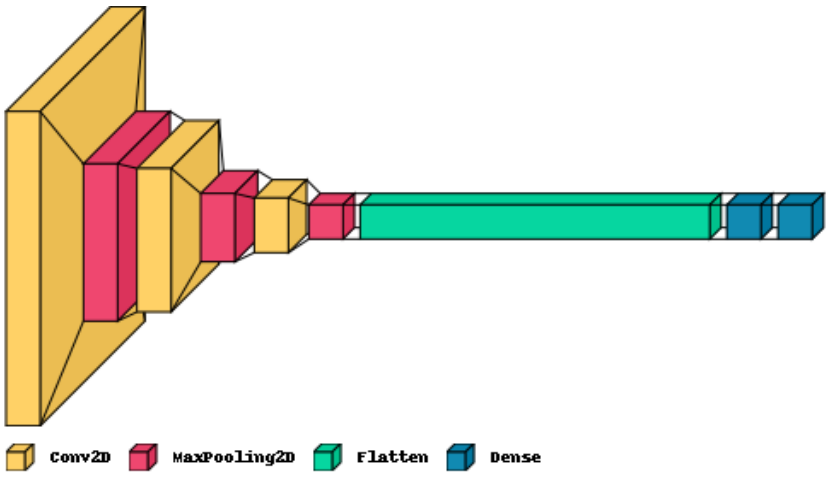
Το δίκτυο αυτό, καθώς και όλες οι παραλλαγές του, δημιουργήθηκαν και εκπαιδεύθηκαν με την βοήθεια των βιβλιοθηκών Python Tensorflow [88] και Keras [89] [1].

Όλα τα μοντέλα εκπαιδεύθηκαν χρησιμοποιώντας Categorical Cross Entropy Loss ως συνάρτηση κόστους και τον Adam Optimizer [99] ως βελτιστοποιητή κατά την διάρκεια της εκπαίδευσης. Ο Adam Optimizer, ή αλλιώς Adaptive Moment Estimation, βελτιστοποιεί τον αλγόριθμο gradient descent συνδυάζοντας δύο αλγορίθμους βελτιστοποίησης: τον αλγόριθμο Momentum [100] και τον Root Mean Square Propagation (RMSP ή RMSprop).

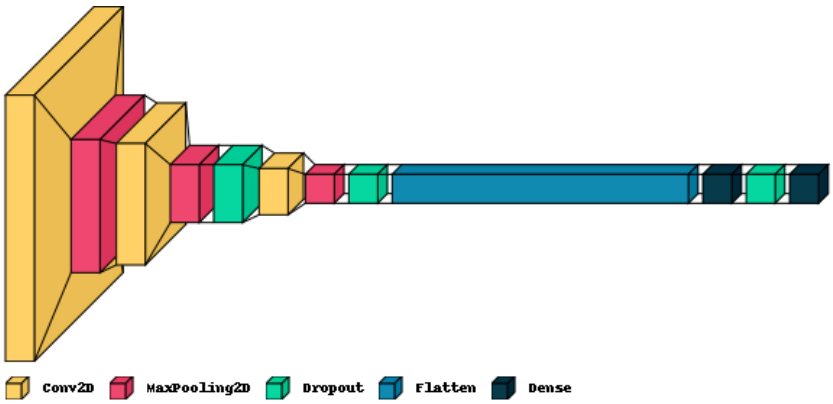
[1] Για την παραγωγή των εικόνων 6.6 και 6.7 χρησιμοποιήθηκε το VisualKeras [98]



Σχήμα 6.5: Αρχιτεκτονική Βασικού Μοντέλου Βαθιάς Μάθησης



Σχήμα 6.6: Τρισδιάστατη απεικόνιση αρχιτεκτονικής βασικού μοντέλου βαθιάς μάθησης



Σχήμα 6.7: Τρισδιάστατη απεικόνιση αρχιτεκτονικής βασικού μοντέλου βαθιάς μάθησης με dropout

Πίνακας 6.1: ReduceLROnPlateau

Fixed Parameter	**Value**
monitor	val_loss
factor	0.1
min_lr	0.0001

Αν και το βασικό μοντέλο έδειξε να έχει σχετικά ικανοποιητικά αποτελέσματα [2], για την βελτίωσή του δοκιμάστηκαν διάφορες τροποποιήσεις αλλά και τεχνικές κανονικοποίησης. Αρχικά, εισήχθη ένα στρώμα dropout σε τρία σημεία του μοντέλου. Πιο συγκεκριμένα εισήχθη dropout μεταξύ του δεύτερου στρώματος Max Pooling και τρίτου συνελικτικού, μεταξύ του τελευταίου Max Pooling και του στρώματος flatten, και μεταξύ των δύο πλήρως συνδεδεμένων στρωμάτων στο τέλος. Το dropout εισήχθη σε διάφορα ποσοστά. Το ποσοστό αυτό δηλώνεται ως ένας δεκαδικός αριθμός a (π.χ. 0.1 αντιστοιχεί σε 10 %) και εκφράζει το μέρος των νευρώνων οι οποίοι θα απενεργοποιηθούν, αφήνοντας το υπόλοιπο $1 - a$ να μάθει τα δεδομένα. Επομένως, όσο μεγαλύτερο το ποσοστό του dropout, τόσο μεγαλύτερο και το ποσοστό των νευρώνων που απενεργοποιούνται.

Έπειτα, δοκιμάστηκε η χρήση του Early Stopping για διάφορους αριθμούς patience και του callback "ReduceLROnPlateau". Στο Early Stopping η μετρική που παρακολοθούνταν για να αποφασιστεί το αν θα διακοπεί νωρίς η εκπαίδευση ήταν το validation loss. Στον παρακάτω πίνακα (6.1) παρουσιάζονται οι παράμετροι του ReduceLROnPlateau.

Ακόμη, έγιναν δοκιμές σε διάφορες υπερπαραμέτρους, όπως τα epochs και το batch size. Οι εποχές εκπαίδευσης του δικτύου, ή epochs, αποτελούν μια υπερπαράμετρο που είναι σημαντικό να εξερευνηθεί, εφόσον επηρεάζουν σημαντικά την εκπαιδευτική διαδικασία. Ο αριθμός των εποχών είναι άρρικτα συνδεδεμένος με το overfitting και το underfitting, γιατί με λίγες εποχές το μοντέλο μπορεί να μην έχει αρκετό καιρό να μάθει τα δεδομένα, ενώ με πολλές εποχές το μοντέλο μπορεί να αρχίσει να απομνημονεύει τα δεδομένα εκπαίδευσης αντί να γενικεύει, οδηγώντας σε υπερπροσαρμογή.

Όσον αφορά το batch size, καταρχάς μπορεί να επηρεάσει την υπολογιστική αποδοτικότητα, αφού μικρότεροι αριθμοί batch size μπορούν να οδηγήσουν σε πιο αργή εκπαίδευση. Έπειτα, μπορεί να επηρεάσει την γενίκευση που μπορεί να κάνει ένα μοντέλο, επειδή μεγάλα batch sizes μπορούν να οδηγήσουν σε χαμηλότερη ικανότητα γενίκευσης [101].

Επίσης, πειραματιστήκαμε με διάφορα ποσοστά διαχωρισμού μεταξύ σετ εικπαίδευσης, σετ ελέγχου και σετ επιβεβαίβωσης. Το σετ χωρίστηκε σε 70 % σετ εκπαίδευσης, 15 % σετ ελέγχου και 15 % σετ επιβεβαίωσης, αλλά και σε 80 % σετ

[2]Τα αποτελέσματα θα παρουσιασθούν και θα αναλυθούν στο επόμενο κεφάλαιο

εκπαίδευσης, 10 % σετ ελέγχου και 10 % σετ επιβεβαίωσης. Για την λειτουργία αυτή χρησιμοποιήθηκε η συνάρτηση train_test_split από την βιβλιοθήκη Python Scikit-learn [91].

Είναι σημαντικό να αναφερθεί πως για τον διαχωρισμό των εικόνων επιστρατεύτηκε η τεχνική "stratification" (διαστρωμάτωση). Το stratification εξασφαλίζει πως η κατανομή των κλάσεων μεταξύ του σετ εκπαίδευσης, ελέγχου και επιβεβαίωσης είναι η ίδια. Αυτό βοηθά στην αποφυγή πιθανών ζητημάτων, όπως η εισαγωγή μεροληψίας ή η απόκτηση αναξιόπιστων αποτελεσμάτων λόγω της άνισης κατανομής των κλάσεων μεταξύ των τριών σετ.

Επιπροσθέτως, δοκιμάστηκε ο διαφορετικός διαχωρισμός σε 10 επαναλήψεις. Είναι σημαντικό να αναφερθεί πως ο διαχωρισμός γίνεται κάθε φορά με τρόπο μη τυχαίο και ελέγχεται από έναν αριθμό random state, για λόγους αναπαραγωγιμότητας. Με την ίδια λειτουργία train_test_split, το σετ χωρίστηκε στα προαναφερθέντα ποσοστά, αλλάζοντας σε κάθε επανάληψη το random state για να προκύψουν σετ εικόνων διαφορετικά χωρισμένα. Ύστερα, το μοντέλο εκπαιδεύτηκε χρησιμοποιώντας κάθε σετ από τα 10, αξιολογήθηκε βάσει διαφόρων μετρικών και ύστερα υπολογίσθηκε ο μέσος όρος και η τυπική απόκλισή τους.

Κεφάλαιο 7

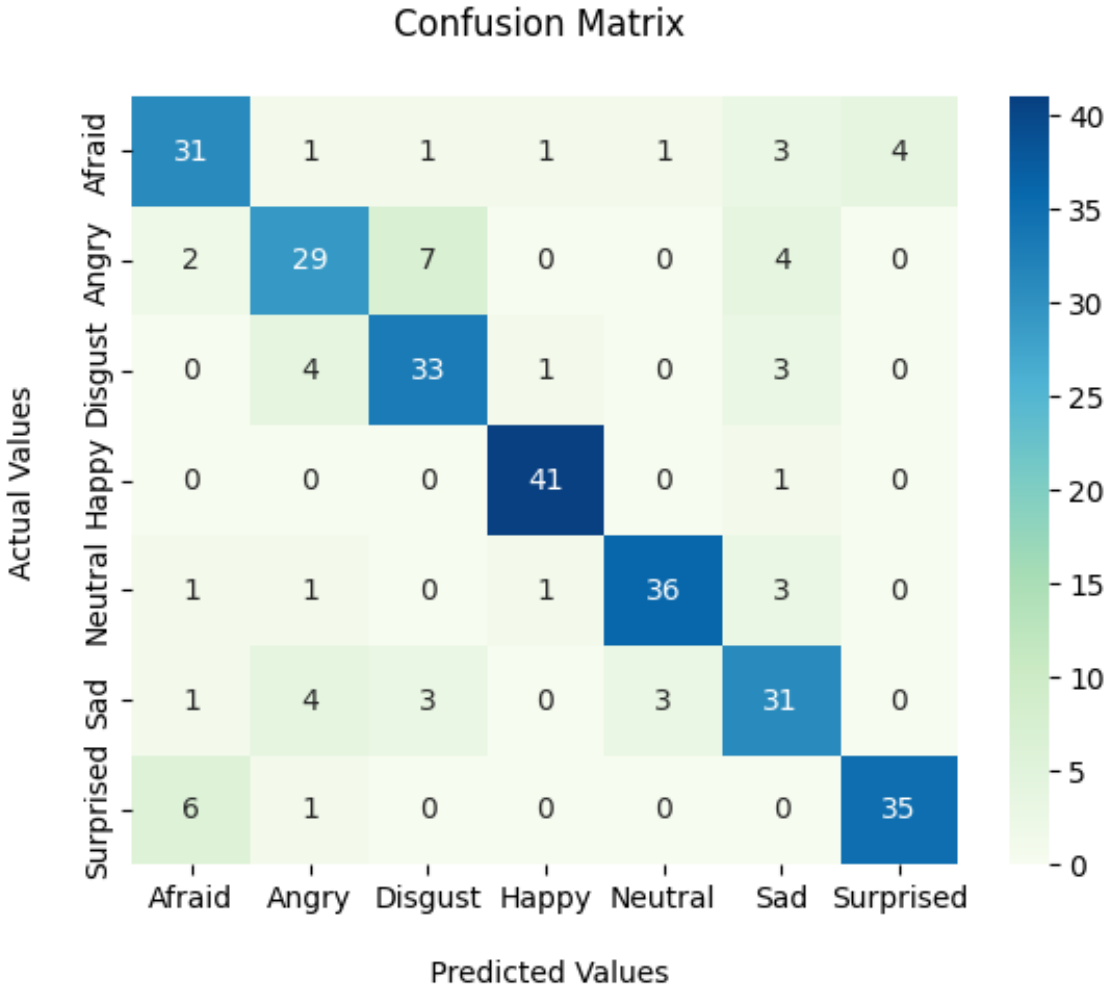
Αποτελέσματα

Το παρόν κεφάλαιο παρέχει μια σφαιρική και συνοπτική επισκόπηση των αποτελεσμάτων που προέκυψαν από την εφαρμογή των μεθόδων και των τεχνικών που περιγράφηκαν στα προηγούμενα κεφάλαια. Σε αυτό το κεφάλαιο, παρουσιάζονται τα ποσοτικά αποτελέσματα των πειραμάτων, καθώς και οι αναλύσεις και οι ερμηνείες που προκύπτουν από αυτά. Επίσης, θα γίνει αναφορά σε ενδεχόμενες παρατηρήσεις, ανακαλύψεις και ευρήματα που προέκυψαν κατά τη διεξαγωγή των πειραμάτων. Η σύνδεση των αποτελεσμάτων με τις στρατηγικές και τους στόχους που καθορίστηκαν στην αρχή της εργασίας θα αποτελέσει τη βάση για την ενδελεχή αξιολόγηση και ερμηνεία των παρουσιαζόμενων αποτελεσμάτων. Οι αποφάσεις που λήφθηκαν, οι επιδόσεις του μοντέλου σε σχέση με τους στόχους της εργασίας και οι πιθανές προκλήσεις θα συζητηθούν ενδελεχώς με σκοπό τη σύνθεση μιας ολοκληρωμένης εικόνας για την επίτευξη των στόχων της παρούσας ερευνητικής εργασίας.

7.1 Αποτελέσματα - Μετρικές Αξιολόγησης

Αρχικά, το σετ KDEF με τις εικόνες τραβηγμένες υπό γωνία (εν συντομία θα αναφέρεται ως KDEF - Full), χωρίσθηκε σε τρία μέρη: 80 % σετ εκπαίδευσης, 10 % σετ ελέγχου και 10 % σετ επιβεβαίωσης. Το βασικό μοντέλο στην μορφή στην οποία παρουσιάσθηκε στο προηγούμενο κεφάλαιο, εκπαιδεύθηκε αρχικά με 50 epochs και batch size 24, και απέφερε ορθότητα 80.54 %. Ο πίνακας σύγχυσης, η γραφική Receiver Operating Characteristic (ROC) και η γραφική Precision-Recall παρουσιάζονται στις εικόνες 7.1 και 7.2.

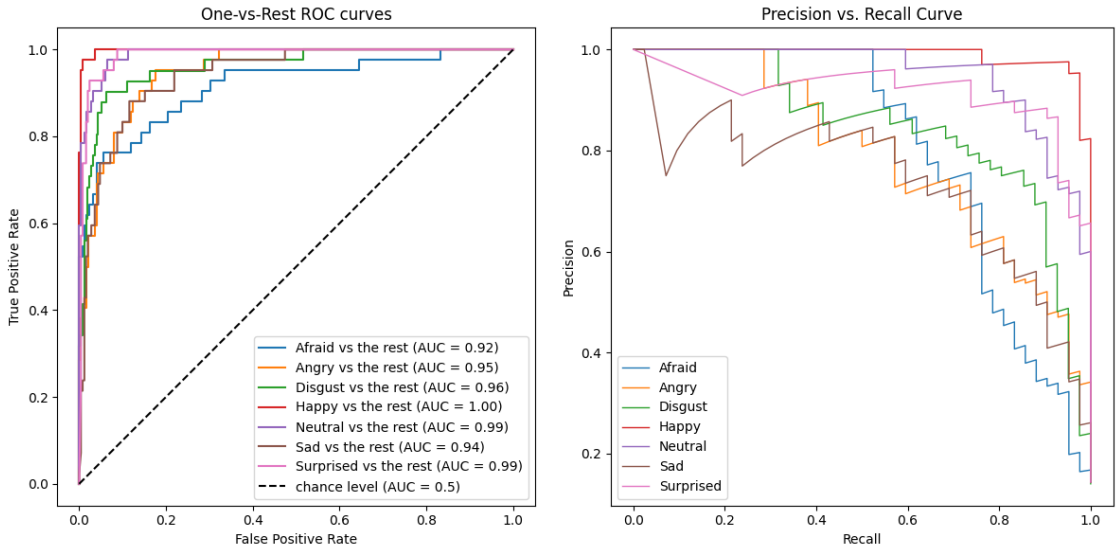
Σύμφωνα με τον πίνακα σύγχυσης, το μοντέλο φαίνεται να μπερδεύει αρκετές εικόνες φόβου με όλα τα άλλα συναισθήματα, αλλά ιδίως με το συναίσθημα της λύπης και της έκπληξης. Το ίδιο συμβαίνει και με την λύπη. Δηλαδή εικόνες που απεικονίζουν θυμό, φόβο, αηδία, ή το ουδέτερο συναίσθημα προβλέπονται ως λύπη. Δύο άλλα συναισθήματα που φαίνεται να συγχέει το μοντέλο είναι ο θυμός και η αηδία. Τέσσερις εικόνες που απεικόνιζαν αηδία προβλέφθηκαν ως θυμός, και επτά εικόνες το αντίστροφο. Αυτό θα μπορούσε να δικαιολογηθεί αν αναλογιστεί κανείς πως τα



Σχήμα 7.1: Πίνακας σύγχυσης βασικού μοντέλου για το KDEF (Full)

δύο αυτά συναισθήματα εκδηλώνονται με παρόμοιες εκφράσεις του προσώπου. Σύμφωνα με το ROC Curve (εικόνα 7.2), όλες οι καμπύλες πλησιάζουν το σημείο (0, 1) το οποίο εκφράζει μεγάλη ευαισθησία (sensitivity) και μικρό False Positive Rate. Σύμφωνα με την μέτρηση Area Under the Curve (AUC), το πιο επιτυχές συναίσθημα ήταν η χαρά, όπως άλλωστε φαίνεται και στον πίνακα σύχυσης. Επίσης, ξεπερνούν κατά πολύ την καμπύλη της τυχαίας ταξινόμησης με το μαύρο χρώμα. Η καμπύλη αυτή εκφράζει την καμπύλη ROC την οποία θα είχε ένας ταξινομητής ο οποίος μαντεύει στην τύχη.

Όσον αφορά το σετ KDEF που περιέχει μόνο τις μπροστινές εικόνες (εν συντομία θα αναφέρεται ως KDEF (Front)), το ίδιο μοντέλο απέφερε 88.77 % ορθότητα. Τα αντίστοιχα αποτελέσματα παρουσιάζονται στην εικόνα 7.3 και 7.4. Σε αυτή την περίπτωση η χαρά προβλέπεται επιτυχώς. Ο θυμός, ωστόσο, πάλι συγχέεται με την αηδία αλλά και με την λύπη, και παράλληλα μία εικόνα που εκφράζει αηδία προβλέφθηκε ως θυμός. Ακόμη, το μοντέλο αυτό έδειξε να μπερδεύει για άλλη μια φορά την αηδία με κάποια αρνητικά συναισθήματα όπως ο θυμός και η λύπη. Τέλος,



Σχήμα 7.2: Αποτελέσματα βασικού μοντέλου για το KDEF (Full)

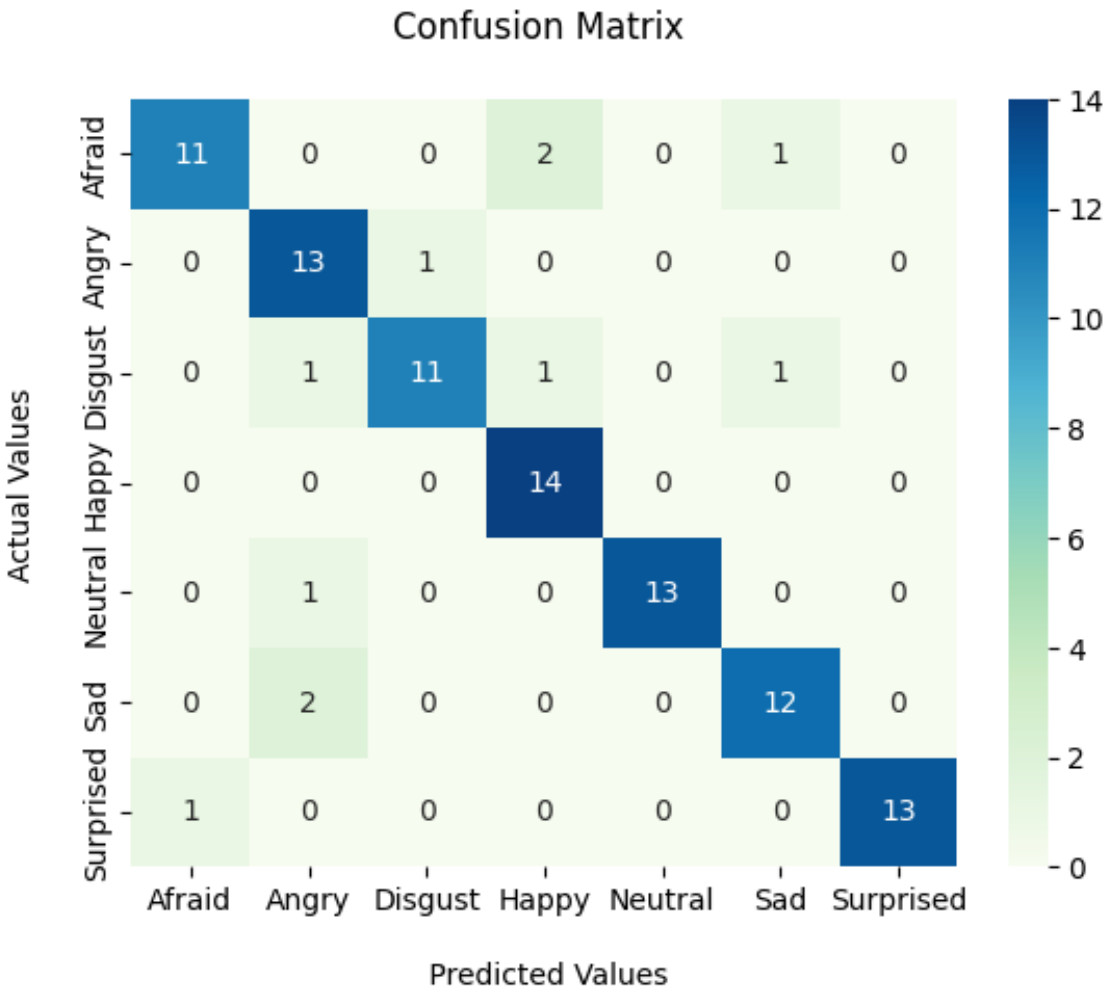
η έκπληξη προβλέφθηκε με σχετικά μεγάλη επιτυχία, αφού μόνο μία εικόνα ήταν ψευδώς αρνητική και προβλέφθηκε ως θυμός.

Σύμφωνα με τις καμπύλες ROC (αριστερά στην εικόνα 7.4), το κάθε συναίσθημα ξεχωριστά προβλέπεται επιτυχώς, ιδίως αυτά με την μεγαλύτερη τιμή στην Area Under the Curve (AUC), όπως η χαρά, η έκπληξη και το ουδέτερο συναίσθημα. Ακόμη, σύμφωνα με την καμπύλη Precision Recall (δεξιά στην εικόνα 7.4), υπάρχει πολύ καλή ισορροπία μεταξύ της ανάκλησης και της ακρίβειας για πολλά συναισθήματα, ιδίως για αυτά που τείνουν προς το σημείο $(1, 1)$ όπου η ισορροπία των δύο είναι η καλύτερη.

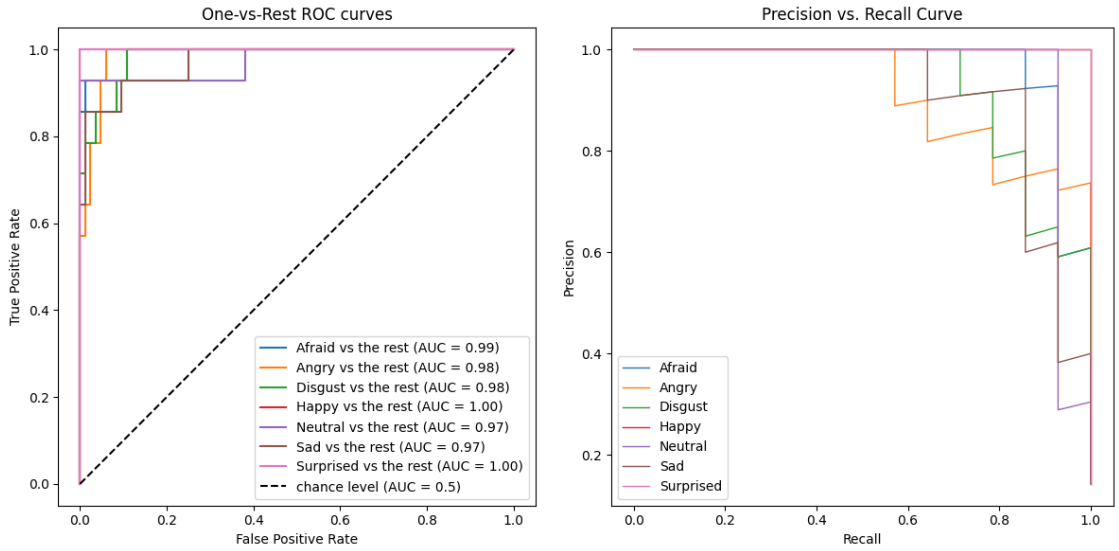
Όπως φαίνεται, η μεγαλύτερη ορθότητα επιτυγχάνεται χρησιμοποιώντας μόνο τις μπροστινές φωτογραφίες. Αυτό υποδεικνύει πως είναι πιθανόν πιο δύσκολη η αναγνώριση των συναισθημάτων στις εικόνες με πλαϊνή γωνία λήψης.

Ακόμη, εφαρμόσθηκε το ίδιο ακριβώς μοντέλο στο σετ JAFFE, το οποίο απέφερε ορθότητα 81 %. Ο πίνακας σύγχυσης, η γραφική Receiver Operating Characteristic (ROC) και η γραφική Precision-Recall παρουσιάζονται στις εικόνες 7.5 και 7.6. Σύμφωνα με τον πίνακα σύγχυσης, ο φόβος, η χαρά, η λύπη και η έκπληξη προβλέφθηκαν με απόλυτη επιτυχία. Αντίθετα, ο θυμός και το ουδέτερο συναίσθημα είχαν από μία εικόνα η οποία προβλέφθηκε σωστά, ενώ η αηδία φάνηκε να είναι το πιο δύσκολο συναίσθημα για να προβλεφθεί, με μόνο μια σωστή πρόβλεψη. Οι καμπύλες ROC και Precision Recall επιβεβαιώνουν τα παραπάνω ευρήματα.

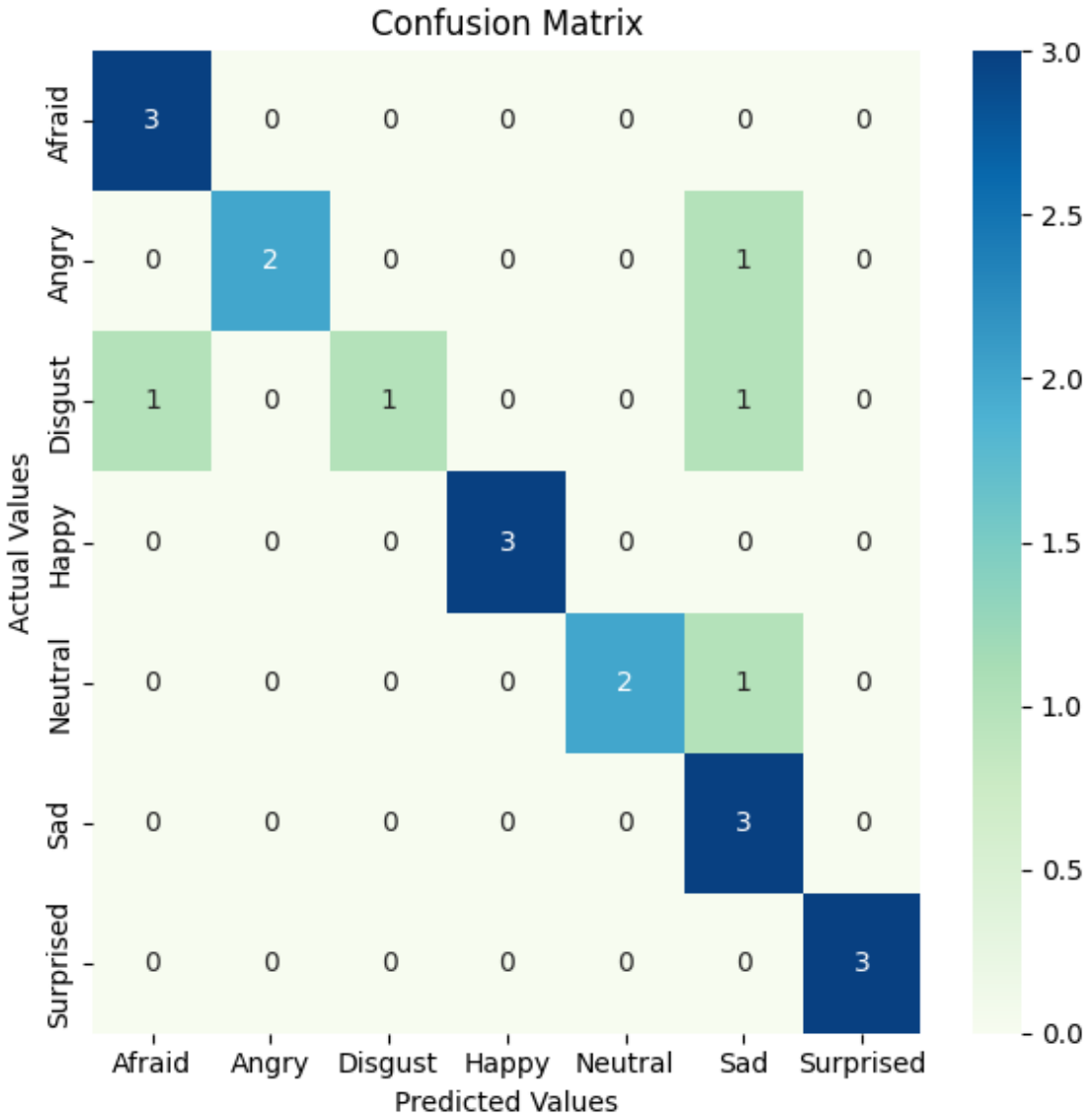
Επιπροσθέτως, η μέση ακρίβεια (precision) έφτασε το 0.893, ενώ η μέση ανάκληση (recall) έφτασε το 0.81. Υπολογίζοντας τον αρμονικό τους μέσο παίρνουμε το F1-score το οποίο ήταν 0.803. Η τιμή για το Cohen's kappa, από την άλλη, ήταν 0.777 ενώ το Area Under the Curve (AUC) για το Reiceiver Operating Characteristic Curve (ROC) ήταν 0.9841.



Σχήμα 7.3: Πίνακας σύγχυσης βασικού μοντέλου για το KDEF (Front)



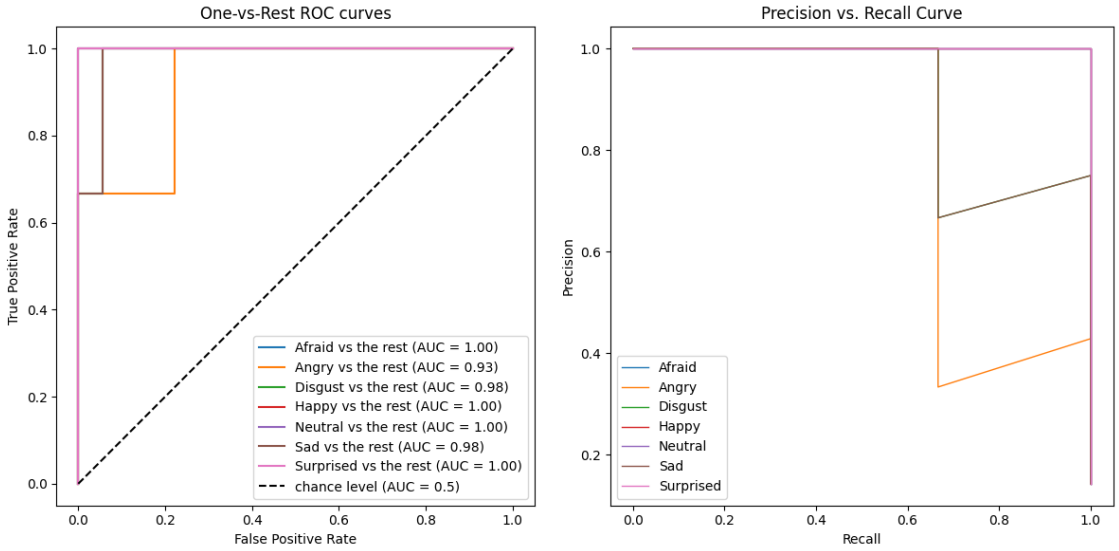
Σχήμα 7.4: Αποτελέσματα βασικού μοντέλου για το KDEF (Front)



Σχήμα 7.5: Πίνακας σύγχυσης βασικού μοντέλου για το JAFFE

Πίνακας 7.1: Σύμπτυξη αποτελεσμάτων βασικού μοντέλου

	Evaluation metrics					
Model	**Accuracy**	**Precision**	**Recall**	**F1 score**	**Cohen's Kappa**	**ROC AUC**
KDEF (Front)	88.77 %	0.897	0.888	0.888	0.869	0.9864
KDEF (Full)	80.54 %	0.807	0.805	0.806	0.773	0.9632
JAFFE	81 %	0.893	0.810	0.803	0.777	0.9841



Σχήμα 7.6: Αποτελέσματα βασικού μοντέλου για το JAFFE

Πίνακας 7.2: Πιθανές τιμές υπερπαραμέτρων

Υπερπαράμετρος	**Τιμές**
Epochs	30, 50, 70, 90
Dropout Rate	0.1, 0.2, 0.3
Patience (Early Stopping)	5, 10, 20
Patience (ReduceLROnPlateau)	5, 10, 20

Ως επόμενο βήμα, εφαρμόσαμε διάφορες τροποποιήσεις στο βασικό μοντέλο και παρατηρήσαμε τα αποτελέσματα. Οι βασικές υπερπαράμετροι οι οποίες μεταβλήθηκαν ήταν τα epochs, το dropout rate, το patience για το Early Stopping και το patience για το ReduceLROnPlateau. Επίσης, εφαρμόσθηκαν ξεχωριστά αλλά και συνδυαστικά το Early Stopping και το ReduceLROnPlateau [1]. Οι τιμές που δοκιμάσθηκαν βρίσκονται στον πίνακα 7.2 και τα αναλυτικά αποτελέσματα στο παράρτημα στο τέλος της εργασίας.

Στον πίνακα 7.3 παρατίθενται οι καλύτερες υπερπαράμετροι για το κάθε μοντέλο ξεχωριστά.

7.2 Αποτελέσματα Διαφορετικών Διαχωρισμών

Για το καλύτερο μοντέλο κάθε σετ, δοκιμάστηκαν 10 διαφορετικοί διαχωρισμοί train_test_split για να ελεγχθεί πώς αποδίδει σε διαφορετικές εικόνες. Με βάση το καλύτερο μοντέλο, πειραματιστήκαμε λίγο παραπάνω με το patience και το batch size με σκοπό να παρατηρήσουμε αν τα αποτελέσματα βελτιώνονται παραπάνω. Τα

[1] Σε κάθε επανάληψη, στο Early Stopping και το ReduceLROnPlateau χρησιμοποιήθηκαν οι ίδιες τιμές patience

Πίνακας 7.3: Καλύτερες παράμετροι μετά από τα πειράματα

	Parameters				
Set	**Dropout**	**Epochs**	**Patience**	**Early Stopping**	**Reduce LR**
KDEF (Full)	0.3	90	10	Όχι	Ναι
KDEF (Front)	0.3	70	5	Ναι	Όχι
JAFFE	0.2	50	10	Όχι	Ναι

αποτελέσματα παρουσιάζονται αναλυτικά στον πίνακα 7.4.

Όσον αφορά το πλήρες KDEF, η αύξηση του batch size επηρέασε θετικά την ορθότητα. Πιο συγκεκριμένα, η μέση ορθότητα των 10 επαναλήψεων ήταν 83.69 %, με αρκετά μικρή τυπική απόκλιση. Αυτό σημαίνει πως η ορθότητα παραμένει σχετικά σταθερή σε κάθε επανάληψη, επομένως το μοντέλο έχει την ικανότητα να γενικεύσει και να προβλέψει επιτυχώς διαφορετικές εικόνες. Ακόμη, ο μέσος όρος (0.8343) και η τυπική απόκλιση (0.0240) του F1-score (δηλαδή ο αρμονικός μέσος ακρίβειας και ανάκλησης), υποδεικνύουν πως το μοντέλο έχει καλές επιδόσεις τόσο όσον αφορά τον εντοπισμό θετικών περιπτώσεων όσο και την αποφυγή ψευδώς θετικών.

Πρέπει να σημειωθεί πως, εφόσον έχουμε multi-class classification, το F1-score της κάθε επανάληψης υπολογίζεται ως ο μέσος όρος των F1-score της κάθε κλάσης, δηλαδή του κάθε συναισθήματος. Ο λόγος πίσω από αυτό είναι πως η κάθε μετρική υπολογίζεται με τον "One-Vs-All" τρόπο, δηλαδή για κάθε συναίσθημα ως θετική κλάση θεωρείται το συναίσθημα αυτό και ως αρνητική κλάση τα υπόλοιπα 6. Αυτό οδηγεί στον υπολογισμό επτά F1-score, των οποίων βρίσκεται ο μέσος όρος.

Προχωρώντας στο KDEF με τις μπροστινές εικόνες, η μέση ορθότητα του συγκεκριμένου μοντέλου για 10 επαναλήψεις ήταν 86.73 %, με μια αρκετά μικρή τυπική απόκλιση 0.0449. Επομένως, σε αυτή την περίπτωση η ορθότητα παρουσιάζει ελαφρώς μεγαλύτερες διακυμάνσεις ανά επανάληψη, σε σχέση με το προηγούμενο σετ εικόνων. Επιπροσθέτως, ο μέσος όρος του F1-score ήταν 0.8650 και η τυπική του απόκλιση 0.0454, το οποίο σημαίνει πως το μοντέλο μπορεί να προβλέψει τις "θετικές" περιπτώσεις με αρκετή ευκολία.

Τέλος, σχετικά με το JAFFE, η μέση ορθότητα των 10 επαναλήψεων άγγιξε το 88.09 % με τυπική απόκλιση 0.0648, το οποίο υποδεικνύει πάλι ελαφρώς μεγαλύτερες διακυμάνσεις στην ορθότητα ανάλογα με τον διαχωρισμό του σετ εικόνων σε σετ εκπαίδευσης, ελέγχου και επιβεβαίωσης. Το μέσο F1-score πήρε την τιμή 0.8746 με τυπική απόκλιση 0.0710, πράγμα που σημαίνει πως διακρίνει με ευκολία τις θετικές περιπτώσεις, αλλά με κάποια ψευδώς θετικά δείγματα.

Πίνακας 7.4: Αποτελέσματα μετά από 10 επαναλήψεις διαφορετικών διαχωρισμών

		Dataset		
		KDEF (Full)	**JAFFE**	**KDEF (Front)**
Parameters	**Dropout**	0.3	0.1	0.3 (0.5 στο τελευταίο στρώμα)
	Epochs	90	70	70
	Patience (Reduce LR)	15	20	20
	Batch size	48	24	24
Metrics	**Mean Accuracy**	0.8369 ± 0.0231	0.8809 ± 0.0648	0.8673 ± 0.0449
	Average F1	0.8343 ± 0.0240	0.8746 ± 0.0710	0.8650 ± 0.0454
	Average Cohen's Kappa	0.8097 ± 0.0269	0.8611 ± 0.0756	0.8452 ± 0.0524
	Average ROC AUC	0.9747 ± 0.0064	0.9828 ± 0.0135	0.9842 ± 0.0076

7.3 Σύγκριση με υπάρχουσες μεθόδους

Σε αυτό το υποκεφάλαιο τα καλύτερα μοντέλα μας συγκρίνονται με διάφορα μοντέλα που δημιουργήθηκαν για τον σκοπό της ταξινόμησης συναισθημάτων, τα οποία εκπαιδεύθηκαν και ελέγχθηκαν με τα dataset KDEF και JAFFE. Τα αποτελέσματα παρουσιάζονται παρακάτω στους πίνακες 7.5 και 7.6.

Οι Jammoussi et al. [102] χρησιμοποίησαν transfer learning και το προεκπαιδευμένο δίκτυο Alexnet [22] για να εξάγουν τα χαρακτηριστικά των προσώπων και ύστερα τον αλγόριθμο Extreme Learning Machine [103] για προβλέψουν τα επτά συναισθήματα με ορθότητα 81.92 % στο JAFFE και 85.9 % στο KDEF.

Οι Zhou et al. [104] χρησιμοποίησαν επίσης transfer learning με το Alexnet [22], επιστρατεύοντας παράλληλα και τεχνικές εξαγωγής χαρακτηριστικών. Η μέθοδός τους απέφερε 86.43 % ορθότητα στο KDEF.

Οι Zalvarez et al. [105] επιστράτευσαν το προεκπαιδευμένο δίκτυο VGG [21], το οποίο και προσάρμοσαν σε διάφορα dataset εικόνων, συμπεριλαμβανομένων των KDEF και JAFFE. Η μέθοδός τους πέτυχε ορθότητα 72.55 % με τυπική απόκλιση 0.72 στο KDEF, ενώ στο JAFFE κατάφεραν μόλις 49.62 % ορθότητα με 3.71 τυπική απόκλιση.

Οι Sari et al. [106], από την άλλη, χρησιμοποίησαν ένα CNN δύο συνελικτικών

στρωμάτων, σε αντίθεση με το δικό μας μοντέλο που είχε τρία, τα οποία ακολουθούνταν από στρώματα max pooling και 20 % dropout. Παρόλο που, σε γενικές γραμμές, η αρχιτεκτονική ήταν παρόμοια, το μοντέλο τους απέφερε 86.24 % ορθότητα στο KDEF με μπροστινές εικόνες και 82.38 % στο JAFFE.

Οι Shan et al. [107] ακολούθησαν επίσης την τακτική των δύο συνελικτικών στρωμάτων και δύο στρωμάτων υποδειγματολειψίας (subsampling layers) και πέτυχαν μέγιστη ορθότητα 76.74 % στο JAFFE.

Η μέθοδος των Lopes et al. [51] περιλαμβάνει επίσης ένα CNN δύο συνελικτικών στρωμάτων, με μεγέθη φίλτρου $5x5$ και $3x3$ αντίστοιχα. Απέφερε 84.48 % ορθότητα (με τυπική απόκλιση 5 %) στο JAFFE με 6 συναισθήματα και 86.74 % (με τυπική απόκλιση 3 %) στο JAFFE με 7 συναισθήματα. Η μέθοδος των Lopes et al. [51] έχει επίσης αναφερθεί αναλυτικά και στο κεφάλαιο 3. Το μοντέλο των 6 εκφράσεων περιέχει όλα τα συναισθήματα με εξαίρεση το ουδέτερο.

Οι Melaugh et al. [108] πρότειναν επίσης ένα CNN, αυτή τη φορά ενός μόνο συνελικτικού στρώματος. Η κύρια διαφορά με την παραπάνω μέθοδο αλλά και την δική μας ήταν πως διαίρεσαν την εικόνα του προσώπου στην μέση και εκπαίδευσαν μοντέλα και για αυτές τις εικόνες. Η ορθότητα που απέφεραν στο KDEF χρησιμοποιώντας ολόκληρα τα πρόσωπα ήταν 89.4 %, ωστόσο στο JAFFE η ορθότητα άγγιξε μόλις το 76.56 %.

Οι Liew et al. [109] κατέφυγαν σε μεθόδους που δεν χρησιμοποιούν νευρωνικά δίκτυα, χρησιμοποιώντας φίλτρα Gabor, Histograms of Oriented Gradients (HOG) και Fern Feature Discriptors [110] για την εξαγωγή χαρακτηριστικών και ταξινομητές SVM για την εκπαίδευση και πρόβλεψη των συναισθημάτων. Η μέθοδός τους εφαρμόσθηκε στο KDEF και απέφερε 87.2 % προβλέποντας και τα 7 συναισθήματα.

Οι Eng et al. [111] χρησιμοποίησαν επίσης Histograms of Oriented Gradients (HOG) για την εξαγωγή των χαρακτηριστικών και ταξινομητή SVM για την πρόβλεψη των επτά συναισθημάτων. Η παραπάνω μέθοδος απέφερε ορθότητα 76.19 % στο JAFFE και 80.95 % ορθότητα στο KDEF.

7.4 Ερμηνευσιμότητα (Explainability)

Για να κατανοηθούν καλύτερα οι διεργασίες λήψης αποφάσεων των νευρωνικών δικτύων και να παρέχουμε οπτικές εξηγήσεις για τα συναισθήματα που αναγνωρίζονται από τα μοντέλα, εφαρμόσθηκε η τεχνική Gradient-weighted Class Activation Mapping (Grad-CAM) σε κάθε ένα από τα συναισθήματα που εξετάζονται. Το Grad-CAM επιτρέπει να οπτικοποιήσουμε ποιες περιοχές των εικόνων έπαιξαν κρίσιμο ρόλο στις προβλέψεις του μοντέλου.

Δημιουργώντας χάρτες ενεργοποίησης για κάθε συναίσθημα, μπορούμε να εντοπίσουμε τα συγκεκριμένα χαρακτηριστικά του προσώπου, όπως τα μάτια, το στόμα ή

Πίνακας 7.5: Σύγκριση αποτελεσμάτων για το JAFFE

		Metrics
Method	**# Emotions**	**Accuracy**
Jammoussi et al. [102]	7	81.92%
Sari et al. [106]	7	86.24%
Melaugh et al. [108]	7	76.56%
Shan et al. [107]	7	76.7%
Lopes et al. [51]	7	86.74% ± 0.03
Eng et al. [111].	7	76.19%
Our CNN	7	88.09 % ± 0.0648

Πίνακας 7.6: Σύγκριση αποτελεσμάτων για το KDEF (Front)

		Metrics
Method	**# Emotions**	**Accuracy**
Zavarez et al. [105]	7	72.55 % ± 0.72
Eng et al. [111]	7	80.95 %
Sari et al. [106]	7	82.38%
Jammoussi et al. [102]	7	85.9%
Zhou et al. [104]	7	86.43%
Melaugh et al. [108]	7	89.4%
Liew et al. [109]	7	87.2% ± 0.03
Our CNN	7	88.77 %

τα φρύδια, στα οποία το μοντέλο επικεντρώθηκε κατά τη διάρκεια των ταξινομήσεων του. Αυτό δεν αυξάνει μόνο την ερμηνευσιμότητα της απόδοσης του μοντέλου, αλλά βοηθά επίσης στην κατανόηση των βασικών οπτικών σημάτων που συμβάλλουν στην αναγνώριση των συναισθημάτων.

Η μέθοδος αυτή για να λειτουργήσει χρειάζεται ως είσοδο ένα εκπαιδευμένο μοντέλο, το οποίο και χρησιμοποιεί αργότερα για να παράξει τον θερμικό χάρτη. Έτσι χρησιμοποιήθηκε η καλύτερη εκ των 10 επαναλήψεων του μοντέλου για το KDEF (Full), η οποία απέφερε 84.32 % ορθότητα (κατά μέσο όρο το μοντέλο είχε αποφέρει 83.69 % ορθότητα - βλ. πίνακα 7.4).

Για κάθε κλάση, επιλέχθηκαν δύο εικόνες που ταξινομήθηκαν σωστά και δύο εικόνες που ταξινομήθηκαν λάθος και ύστερα εφαρμόσθηκε το Grad-CAM για το τελευταίο συνελικτικό στρώμα. Για παράδειγμα, για το συναίσθημα της λύπης, επιλέχθηκαν στην τύχη δύο εικόνες του KDEF (Full) οι οποίες προβλέφθηκαν επιτυχώς ως λύπη από το μοντέλο, καθώς και δύο εικόνες στις οποίες το μοντέλο αστόχησε και προέβλεψε κάποιο άλλο συναίσθημα. Τα αποτελέσματα φαίνονται στην εικόνα 7.7. Κάθε σειρά εικόνων παρουσιάζει εικόνες ενός μόνο συναισθήματος. Οι δύο πρώτες στήλες αποτελούν εικόνες της κάθε κλάσης που προβλέφθηκαν ορθώς, και οι άλλες δύο στήλες παρουσιάζουν εικόνες της κλάσης που η πρόβλεψη τους δεν ήταν επιτυχής.

Στις εικόνες του συναισθήματος του φόβου οι οποίες προβλέφθηκαν σωστά φαίνεται πως η περιοχή του μετώπου είναι η πιο σημαντική, όπως επίσης και το πηγούνι και μια μικρή περιοχή των ζυγωματικών. Από την άλλη, οι εικόνες του φόβου που δεν ταξινομήθηκαν ορθώς φάνηκε να βρίσκουν σημαντικά μόνο τα ζυγωματικά και το πηγούνι στην μία εικόνα (αριστερά) και μόνο το μέτωπο στην άλλη εικόνα (δεξιά). Πιο συγκεκριμένα, η αριστερή εικόνα προβλέφθηκε ως "αηδία", όπου, όπως διακρίνεται πιο κάτω στην εικόνα, φαίνεται να ενεργοποιεί παρόμοιες περιοχές του προσώπου με τις ορθώς ταξινομημένες εικόνες του συναισθήματος αηδίας.

Όσον αφορά τις εικόνες του συναισθήματος του θυμού, το Grad-CAM στις δύο σωστές προβλέψεις υποδεικνύει ότι τα φρύδια και το άνω χείλος είναι αρκετά σημαντικά για την πρόβλεψη, όπως και το μέτωπο το οποίο κρίνεται ακόμα πιο σημαντικό αφού επισημαίνεται με κόκκινο χρώμα. Αυτές οι περιοχές είναι λογικό να κρίνονται ως σημαντικές, αφού συχνά ο θυμός εκφράζεται με συνοφρυωμένο βλέμμα και πιεσμένα χείλη. Απεναντίας, στις δύο λάθος προβλέψεις φαίνεται πως, αν και διατηρούνται κάποιες περιοχές σημαντικότητας όπως το άνω χείλος και τα φρύδια (αριστερή εικόνα), σε καμία εικόνα δεν επισημάνεται το μέτωπο, γι' αυτό και το μοντέλο μπερδεύεται και προβλέπει λύπη και αηδία.

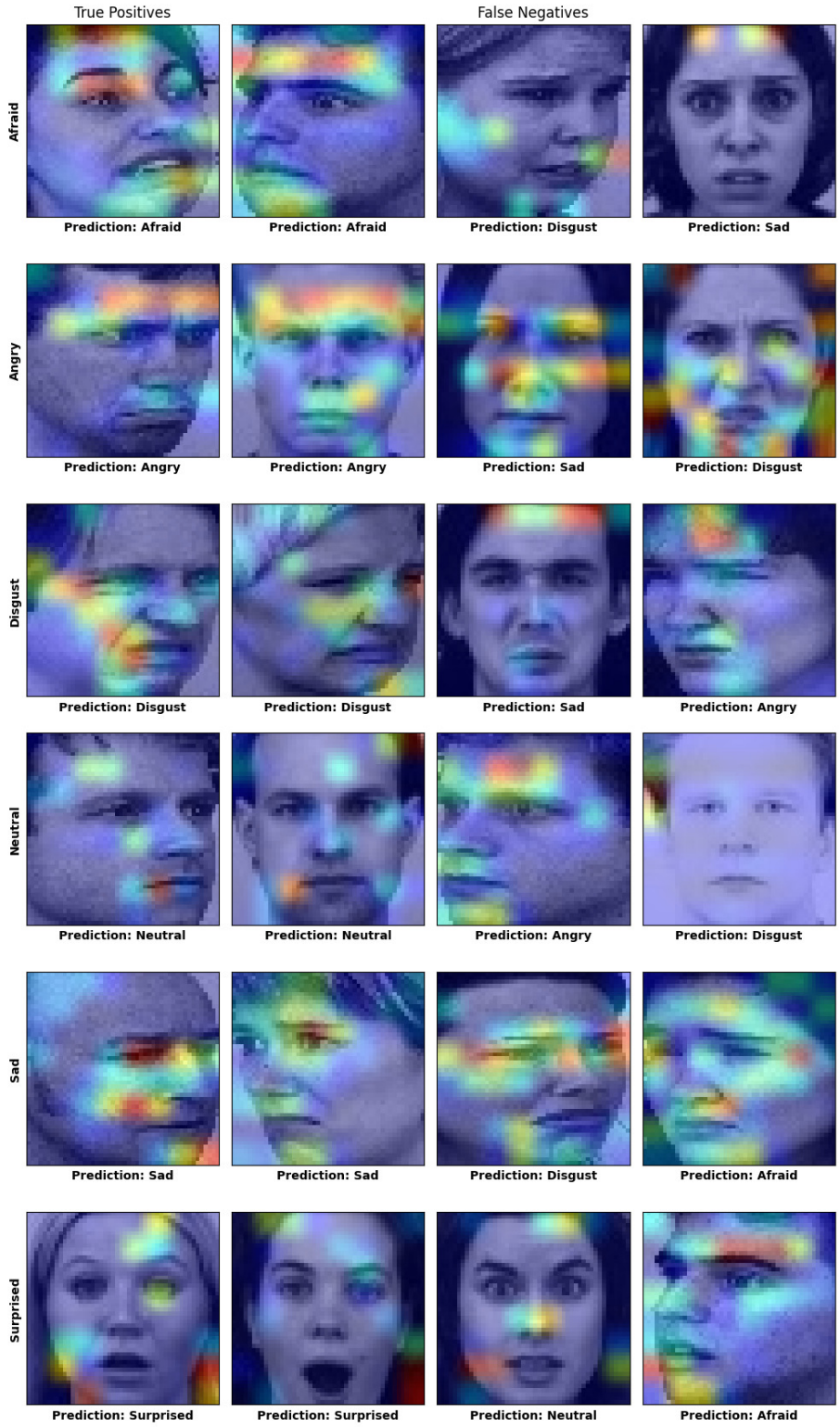
Προχωρώντας στο συναίσθημα της αηδίας, οι εικόνες που προβλέφθηκαν ορθώς μας λένε πως η μύτη, η περιοχή κάτω από τα μάτια, το στόμα και οι κρόταφοι κρίνονται σημαντικά, πράγμα επίσης λογικό, αφού συχνά η απέχθεια εκφράζεται κλείνοντας ελαφρώς τα μάτια και υψώνοντας την μύτη και το στόμα. Αντιθέτως, οι εικόνες που δεν προβλέφθηκαν ως εικόνες που εκφράζουν την αηδία φάνηκε να

δίνουν περισσότερη σημασία στο μέτωπο, παρόλο που έκριναν και την περιοχή γύρω από το στόμα ως σημαντική. Συγκεκριμένα η δεξιά εικόνα φαίνεται να έχει κρίνει ως σημαντικές τις περιοχές του προσώπου που είναι σημαντικές στο συναίσθημα του θυμού, για αυτό και προβλέπεται με αυτή την ετικέτα και όχι ως αηδία.

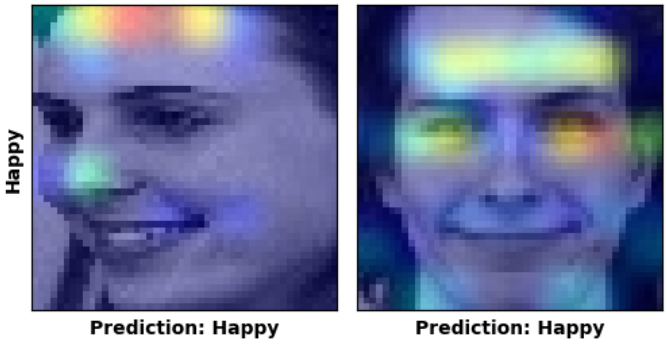
Όσον αφορά το συναίσθημα της χαράς, το μέτωπο, το στόμα και τα ζυγωματικά φάνηκε να έχουν μεγάλη σημαντικότητα στην πρόβλεψη. Σε αυτή την περίπτωση, όμως, όλες οι εικόνες χαράς του σετ ελέγχου προβλέφθηκαν ορθώς, οπότε δυστυχώς δεν υπάρχει σύγκριση, για αυτό τον λόγο και τα αποτελέσματα παρουσιάζονται μεμονωμένα στην εικόνα 7.8. Προχωρώντας στο ουδέτερο συναίσθημα, στις ορθές προβλέψεις σημειώθηκαν ως σημαντικές μικρές περιοχές στην μύτη, τους κροτάφους, το μέτωπο και τις άκρες του στόματος. Αντιθέτως, στις λανθασμένες προβλέψεις είτε δεν σημειώθηκε κάτι πάνω στο πρόσωπο (δεξιά), είτε σημειώθηκε ένα μεγαλύτερο μέρος του μετώπου και της σιαγόνου.

Ακόμη, όσον αφορά το συναίσθημα της λύπης, το κέντρο των ματιών, τα ζυγωματικά και το πηγούνι κρίθηκαν από το Grad-CAM ως πολύ σημαντικά για την σωστή πρόβλεψη. Τουναντίον, οι εικόνες λυπημένων προσώπων που δεν προβλέφθηκαν ορθώς, παρόλο που έκριναν ως σημαντικές παρόμοιες περιοχές του προσώπου, φάνηκαν να δίνουν μεγαλύτερη αξία στην περιοχή γύρω από τα μάτια και όχι το κέντρο τους, όπως και στην ράχη της μύτης και τους κροτάφους. Η αριστερή εικόνα έχει προβλεφθεί ως θυμός αφού έχει κρίνει ως σημαντικό το μέτωπο, τα φρύδια και το άνω χείλος, περιοχές που επισημάνθηκαν ως σημαντικές στις εικόνες που εκφράζουν θυμό.

Τέλος, σχετικά με το συναίσθημα της έκπληξης, ανατέθηκε μεγάλη σημαντικότητα στα μάτια, σε ένα μικρό μέρος του μετώπου, καθώς και ελαφρώς στο σαγόνι. Οι παραπάνω περιοχές δικαιολογούνται, αφού συχνά η έκπληξη δηλώνεται με υψωμένα τα μάτια, ζαρωμένο μέτωπο και ανοικτό στόμα το οποίο προκαλεί ένταση στο σαγόνι. Αντιθέτως, οι εικόνες που λανθασμένα δεν προβλέφθηκαν ως εικόνες που εκφράζουν την έκπληξη έδειξαν πως κρίνεται σημαντικό ένα μεγαλύτερο μέρος του μετώπου και των ζυγωματικών, όσο και της μύτης. Ιδιαίτερα η μύτη, δεν φαίνεται να παίζει κάποιον ιδιαίτερο ρόλο στο συναίσθημα της έκπληξης, οπότε θα μπορούσε να δικαιολογήσει και την λάθος πρόβλεψη της δεξιάς εικόνας. Η δεξιά εικόνα προβλέφθηκε ως θυμός πιθανόν γιατί ενεργοποιεί τις ίδιες περιοχές του προσώπου, δηλαδή το πηγούνι, τα ζυγωματικά και το μέτωπο, το οποίο έτσι κι αλλιώς κρίνεται ως πολύ σημαντικό στο συγκεκριμένο συναίσθημα.



Σχήμα 7.7: Ερμηνεία πρόβλεψης με Grad-CAM



Σχήμα 7.8: Ερμηνεία πρόβλεψης συναισθήματος χαράς με Grad-CAM

Κεφάλαιο 8

Συμπεράσματα και μελλοντική έρευνα

Σε αυτό το κεφάλαιο γίνεται μια σύνοψη των αποτελεσμάτων και εξετάζονται οι πιθανές κατευθύνσεις μελλοντικής έρευνας σε αυτό τον τομέα. Αναφέρονται πιθανές βελτιώσεις και επεκτάσεις της διαδικασίας ταξινόμησης συναισθημάτων από πρόσωπα.

8.1 Συμπεράσματα

Στην διπλωματική αυτή εργασία εξερευνήθηκε το πρόβλημα της ταξινόμησης των συναισθημάτων από εικόνες προσώπου. Η άνοδος της τεχνητής νοημοσύνης και η ολοένα αυξανόμενη χρήση της υπολογιστικής όρασης προκάλεσαν την δημιουργία πολλών εφαρμογών που αξιοποιούν την ικανότητα υπολογιστικών συστημάτων να αναγνωρίζουν και να ερμηνεύουν εικόνες. Οι εφαρμογές αυτές καλύπτουν ένα ευρύ φάσμα τομέων, από την ανίχνευση συναισθημάτων σε προσώπα σε κοινωνικές αλληλεπιδράσεις μέχρι τη βελτιστοποίηση της υγείας και την αυτόνομη οδήγηση. Με την ευρεία χρήση των κοινωνικών μέσων και την αυξανόμενη ανάγκη για ανάλυση της συναισθηματικής κατάστασης ατόμων μέσω των διαδικτυακών επικοινωνιών, η ταξινόμηση των συναισθημάτων από εικόνες προσώπου αποκτάει ολοένα και μεγαλύτερη σημασία. Η εργασία αυτή αποτελεί μια συνεισφορά στον τομέα της υπολογιστικής όρασης και της αναγνώρισης συναισθημάτων, προσφέροντας προηγμένες τεχνικές για την ακριβή και αξιόπιστη ταξινόμηση των συναισθημάτων σε εικόνες προσώπων.

Στα πλαίσια της εργασίας αναπτύχθηκαν συνελικτικά τεχνητά νευρωνικά δίκτυα (CNN) τα οποία κατάφεραν να ταξινομήσουν με αρκετή επιτυχία τα επτά συναισθήματα σε εικόνες προερχόμενες από δύο διαφορετικά σετ εικόνων: το Karolinska Directed Emotional Faces (KDEF) και το Japanese Female Facial Expression Database (JAFFE). Το σετ KDEF έγιναν δύο ειδών πειράματα: ένα χρησιμοποιώντας όλες τις εικόνες και ένα αγνοώντας τις εικόνες υπό γωνία λήψης.

Πιο συγκεκριμένα, τα δύο βασικά μοντέλα απέφεραν 77.47 % ορθότητα στο πλήρες KDEF, 88.77 % στο KDEF χωρίς τις φωτογραφίες τραβηγμένες υπό γωνία και 81 % ορθότητα στο σετ εικόνων JAFFE. Μετά από πολλά πειράματα, η μέση ορθότητα του πλήρους KDEF έφτασε στο 83.69 %, και 88.09 % στο JAFFE.

Τα μοντέλα κατάφεραν να διακρίνουν τα περισσότερα συναισθήματα μεταξύ τους, όμως πολύ συχνά έδειξαν να μην τα καταφέρνουν τελείως σε κάποια αρνητικά συναισθήματα, όπως ο θυμός και η αηδία, ή η έκπληξη και ο φόβος. Το πιο επιτυχές συναίσθημα όλων ήταν η χαρά, η οποία ήταν η πιο ευδιάκριτη από όλα τα άλλα συναισθήματα.

Όσον αφορά τις δύο υποκατηγορίες του KDEF, το σετ μόνο με τις μπροστινές εικόνες (KDEF Front) φάνηκε να είναι πιο εύκολο να προβλεφθεί, το οποίο μας οδηγεί να συμπεράνουμε πως οι εικόνες υπό γωνία είναι πιο δύσκολο να προβλεφθούν, αφού το πρόσωπο δεν φαίνεται ολόκληρο.

Ακόμη, συγκρίνοντας με άλλες μεθόδους που προϋπήρχαν, μπορούμε να συμπεράνουμε πως η πρόβλεψη με λιγότερα συναισθήματα είναι σίγουρα πιο επιτυχής, ιδιαίτερα αφαιρώντας τα συναισθήματα που συγχέονται πολύ μεταξύ τους, όπως ο θυμός ή η αηδία. Επομένως, η πρόβλεψη και των επτά συναισθημάτων καθίσταται ορισμένες φορές απαιτητική.

Τέλος, ύστερα από δοκιμές transfer learning μεταξύ των δύο σετ, μπορούμε να πούμε πως είναι πρόκληση να προβλεφθούν εικόνες του ενός σετ από μοντέλο που έχει εκπαιδευθεί αποκλειστικά με άλλο. Στην περίπτωση των JAFFE και KDEF, οι δύο ομάδες εικόνων ήταν αρκετά διαφορετικές μεταξύ τους, ιδιαίτερα όσον αφορά τα χαρακτηριστικά των προσώπων. Πιο συγκεκριμένα, εφαρμόζοντας το μοντέλο του JAFFE στο KDEF, η απόδοση δεν ήταν βέλτιστη, αφού το μοντέλο είχε εκπαιδευθεί με λιγότερες εικόνες και αποκλειστικά με φωτογραφίες γυναικών. Θεωρούμε, επομένως, πως για να αποδόσει ένα νευρωνικό δίκτυο σωστά θα πρέπει να εκπαιδευθεί με ένα σετ εικόνων ετερογενές και μεγάλο. Να περιέχει, δηλαδή, αρκετές εκφράσεις, αρκετές λήψεις, εικόνες γυναικών και ανδρών εξίσου, και πρόσωπα τραβηγμένα από πολλές γωνίες.

8.2 Μελλοντική έρευνα

Η μελλοντική έρευνα στον τομέα της αναγνώρισης συναισθημάτων από πρόσωπα μπορεί να εστιαστεί σε αρκετές κατευθύνσεις προκειμένου να βελτιωθεί η ακρίβεια και η απόδοση των μοντέλων.

Πρώτον, είναι ουσιώδες να εξετάσουμε τον τομέα της προεπεξεργασίας των εικόνων. Βήματα όπως η αφαίρεση του background της εικόνας ή η εξισορρόπηση των χρωμάτων μπορούν να συνεισφέρουν σε μια καλύτερη απόδοση. Η βελτιωμένη αφαίρεση θορύβου και η βελτιωμένη ενίσχυση των χαρακτηριστικών μπορεί να οδηγήσει σε υψηλότερα αποτελέσματα. Επιπροσθέτως, θα ήταν βοηθητικό να αφαιρεθούν και άλλα χαρακτηριστικά στην εικόνα όπως τα μαλλιά, διότι, όπως φάνηκε και στα απο-

τελέσματα, μπορούν να αποπροσανατολίσουν το μοντέλο και να προβεφθεί το λάθος συναίσθημα.

Μετέπειτα, θα ήταν χρήσιμο να εξετασθούν μοντέλα τα οποία δέχονται ως είσοδο έγχρωμες εικόνες. Ιδιαίτερα για το KDEF το οποίο περιέχει αποκλειστικά έγχρωμες εικόνες, θα ήταν ενδιαφέρον να εξετασθούν μοντέλα έγχρωμων εικόνων και να παρατηρηθεί η απόδοσή τους.

Ακόμη, θα ήταν πολύ χρήσιμη η εφαρμογή αύξησης δεδομένων(data augmentation), αναπαράγοντας τις ίδιες εικόνες με περιστροφές, καθρεφτισμούς, zoom κ.α.. Το data augmentation αυξάνει το μέγεθος του σετ εικόνων, καταλήγοντας συχνά σε πιο ισχυρά μοντέλα που έχουν την ικανότητα να γενικεύουν ακόμα πιο αποδοτικά, και μειώνοντας την πιθανότητα υπερπροσαρμογής στα δεδομένα, ειδικότερα για μικρότερα σετ εικόνων όπως το JAFFE.

Ακόμη, βάσει της βιβλιογραφίας συμπεράναμε πως υπάρχουν και άλλα σετ εικόνων τα οποία θα άξιζε να εξετασθούν, όπως το CK [69], το CK+ [52], το Radboud Faces Database (RaFD) [112] και το FER2013 [50]. Δυστυχώς, για κάποια από αυτά, η πρόσβαση ήταν δύσκολη και περιλάμβανε γραφειοκρατικές διαδικασίες ή πληρωμή, οπότε δεν δοκιμάσθηκαν στα πλαίσια της εργασίας αυτής. Θα ήταν επίσης χρήσιμο να εξετασθούν περαιτέρω σετ εικόνων που περιέχουν εικόνες ατόμων διαφόρων εθνικοτήτων και ηλικιών, καθώς τα χαρακτηριστικά τους διαμορφώνονται διαφορετικά, πράγμα το οποίο επηρεάζει την διαδικασία εξαγωγής χαρακτηριστικών και εκπαίδευσης.

Όσον αφορά την αρχιτεκτονική του νευρωνικού δικτύου, θα ήταν σίγουρα χρήσιμο να δοκιμασθούν και άλλες αρχιτεκτονικές. Η εύρεση της ιδανικής αρχιτεκτονικής που θα πετύχει καλά αποτελέσματα εξαρτάται από πολλά, όπως τον τύπο του προβλήματος και τα δεδομένα. Ωστόσο, υπάρχουν τεχνικές εύρεσης αρχιτεκτονικών Νευρωνικών Δίκτύων. Το πεδίο αυτό ονομάζεται Neural Architecture Search [113] [114] και περιλαμβάνει μεθόδους αυτοματοποίησης σχεδιασμού τεχνητών νευρωνικών δικτύων. Το Neural Architecture Search βρίσκει αρκετές εφαρμογές και έχει χρησιμοποιηθεί για την ανάπτυξη CNN τα οποία διαγιγνώσκουν την άνοια [115].

Μία ακόμη ιδέα θα ήταν η δοκιμή του transfer learning με ισχυρά προ-εκπαιδευμένα μοντέλα, όπως το VGG [21] ή το AlexNet [22]. Η χρήση μοντέλων που έχουν ήδη εκπαιδευθεί με χιλιάδες φωτογραφίες θα μπορούσε δυνητικά να βελτιώσει τα αποτελέσματα, ιδιαίτερα σε πιο μικρά datasets εικόνων.

Επίσης, θα ήταν ενδιαφέρον να δοκιμαστούν και άλλες τεχνικές εξόρυξης χαρακτηριστικών, πέραν του deep learning. Κάποιες από αυτές είναι τα φίλτρα Gabor, τα Histograms of Oriented Gradients (HOG) και τα Local Binary Patterns (LBP). Ωστόσο, παραμένοντας στις τεχνικές deep learning, θα ήταν πολύ ενδιαφέρον να εξερευνηθούν και βαθύτερες αρχιτεκντονικές με περισσότερα συνελικτικά στρώματα.

Τέλος, ένας άλλος στόχος θα ήταν η βαθύτερη εξερεύνηση και ρύθμιση των υπερ-παραμέτρων του νευρωνικού δικτύου.

Παράρτημα Α΄

Ακρωνύμια και συντομογραφίες

AI Artificial Intelligence

AUC Area Under the Curve

NLP Natural Language Processing

CV Computer Vision

FER Facial Emotion Recognition

MSE Mean Squared Error

MAE Mean Absolute Error

MLP Multilayer Perceptron

ROC Receiver Operating Characteristic (Curve)

SVM Support Vector Machine

ΤΝΔ Τεχνητό Νευρωνικό Δίκτυο

Bibliography

[1] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American Scientist*, vol. 89, pp. 344–350, 2001.

[2] I. Mihajlovic, "Everything you ever wanted to know about computer vision: Here's a look why it's so awesome," *Towards Data Science*, 2019. [Online]. Available: https://t.ly/U60Fm

[3] G. Vilone and L. Longo, "Classification of explainable artificial intelligence methods through their output formats," *Machine Learning and Knowledge Extraction*, vol. 3, no. 3, pp. 615–661, 2021. [Online]. Available: https://www.mdpi.com/2504-4990/3/3/32

[4] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multi-task cascaded convolutional networks," *CoRR*, vol. abs/1604.02878, 2016. [Online]. Available: http://arxiv.org/abs/1604.02878

[5] E. D. P. Supervisor, "Facial emotion recognition," EDPS website, 2021. [Online]. Available: https://edps.europa.eu/system/files/2021-05/21-05-26_techdispatch-facial-emotion-recognition_ref_en.pdf

[6] A. Mehrabian, "Some referents and measures of nonverbal behavior," *Behavior Research Methods & Instrumentation*, vol. 1, pp. 203–207, 1968.

[7] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.

[8] H.-C. A. I. Stanford University, "Artificial intelligence definitions," Stanford Website, 2020. [Online]. Available: https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf

[9] A. Turing, "On computable numbers, with an application to the entscheidungsproblem," *Proceedings of the London Mathematical Society*, vol. 42,

no. 1, pp. 230–265, 1936.

[10] A. M. TURING, "I.—Computing Machinery and intelligence," *Mind*, vol. LIX, no. 236, pp. 433–460, 10 1950. [Online]. Available: https://doi.org/10.1093/mind/LIX.236.433

[11] A. Newell and H. Simon, "The logic theory machine–a complex information processing system," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 61–79, 1956.

[12] OpenAI, "Chatgpt: Large-scale language model for conversational ai," OpenAI Website, 2021. [Online]. Available: https://openai.com/research/chatgpt/

[13] J. Fürnkranz, "Decision tree," in *Encyclopedia of Machine Learning and Data Mining*. Springer, 2017. [Online]. Available: https://doi.org/10.1007/978-1-4899-7502-7_1380-1

[14] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001.

[15] L. van der Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008. [Online]. Available: https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

[16] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," 2018, cite arxiv:1802.03426Comment: Reference implementation available at http://github.com/lmcinnes/umap. [Online]. Available: http://arxiv.org/abs/1802.03426

[17] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, "Linear discriminant analysis: A detailed tutorial," *AI Commun.*, vol. 30, pp. 169–190, 2017.

[18] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review*, vol. 65 6, pp. 386–408, 1958.

[19] W. Mcculloch and W. Pitts, "A logical calculus of ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 127–147, 1943.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735–80, 12 1997.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv 1409.1556*, 09 2014.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[23] L. Ilias, I. M. Kazelidis, and D. Askounis, "Multimodal detection of social spambots in twitter using transformers," 2023.

[24] L. Ilias, D. Askounis, and J. Psarras, "Detecting dementia from speech and transcripts using transformers," *Computer Speech Language*, vol. 79, p. 101485, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0885230823000049

[25] ——, "Multimodal detection of epilepsy with deep neural networks," *Expert Systems with Applications*, vol. 213, p. 119010, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417422020280

[26] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Machine Learning Proceedings 1992*, D. Sleeman and P. Edwards, Eds. San Francisco (CA): Morgan Kaufmann, 1992, pp. 249–256. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9781558602472500371

[27] I. Kononenko, E. Šimec, and M. Robnik-Sikonja, "Overcoming the myopia of inductive learning algorithms with relieff," *Applied Intelligence*, vol. 7, pp. 39–55, 01 1997.

[28] M. Robnik-Sikonja and I. Kononenko, "An adaptation of relief for attribute estimation in regression," *ICML '97: Proceedings of the Fourteenth International Conference on Machine Learning*, 02 2000.

[29] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1, pp. 273–324, 1997, relevance. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S000437029700043X

[30] J. H. Moore and B. C. White, "Tuning relieff for genome-wide genetic analysis," in *Evolutionary Computation,Machine Learning and Data Mining in Bioinformatics*, E. Marchiori, J. H. Moore, and J. C. Rajapakse, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 166–175.

[31] C. S. Greene, N. M. Penrod, J. Kiralis, and J. H. Moore, "Spatially uniform relieff (surf) for computationally-efficient filtering of gene-gene interactions," *BioData Min*, vol. 2, no. 1, p. 5, Sep 2009.

[32] G. H. John, R. Kohavi, and K. Pfleger, "Irrelevant features and the subset selection problem," in *Proceedings Fifth International Conference on Machine Learning*. Morgan Kaufmann, 1994, pp. 121–129.

[33] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, pp. 389–422, 01 2002.

[34] K. Hao, "Ai is sending people to jail—and getting it wrong," *MIT Technology Review*, January 2019. [Online]. Available: https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/

[35] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1135–1144. [Online]. Available: https://doi.org/10.1145/2939672.2939778

[36] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[37] L. Ilias, F. Soldner, and B. Kleinberg, "Explainable verbal deception detection using transformers," 2022.

[38] L. Ilias and D. Askounis, "Explainable identification of dementia from transcripts using transformer networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 4153–4164, 2022.

[39] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626.

[40] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 06 2014.

[41] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015.

[42] L. Nwosu, H. Wang, J. Lu, I. Unwala, X. Yang, and T. Zhang, "Deep convo-

lutional neural network for facial expression recognition using facial parts," in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, 2017, pp. 1318–1321.

[43] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, 2001, pp. I–I.

[44] M. J. Lyons, ""Excavating AI" Re-excavated: Debunking a Fallacious Account of the JAFFE Dataset," Jul. 2021, n.b. All JAFFE images in this article are subject to specific terms of use and may not be reused without permission, regardless of the license applied to the document as a whole. [Online]. Available: https://doi.org/10.5281/zenodo.5147170

[45] M. J. Lyons, M. Kamachi, and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets (IVC Special Issue)," Sep. 2020, This manuscript is a modified version of a conference article, that was invited for publication in a special issue of Image and Vision Computing dedicated to a selection of articles from the IEEE Face & Gesture 1998 conference. The special issue never materialized. [Online]. Available: https://doi.org/10.5281/zenodo.4029680

[46] S. P. Khandait, R. C. Thool, and P. D. Khandait, "Automatic facial feature extraction and expression recognition based on neural network," *CoRR*, vol. abs/1204.2073, 2012. [Online]. Available: http://arxiv.org/abs/1204.2073

[47] G. Yolcu, I. Oztel, S. Kazan, C. Oz, K. Palaniappan, T. E. Lever, and F. Bunyak, "Deep learning-based facial expression recognition for monitoring neurological disorders," in *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017, pp. 1652–1657.

[48] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and validation of the radboud faces database," *Cognition and Emotion*, vol. 24, no. 8, pp. 1377–1388, 2010. [Online]. Available: https://doi.org/10.1080/02699930903485076

[49] T. Arora, P. Chaubey, M. S. Raman, B. Kumar, N. Yagnam, P. Anjani, H. Ahmed, A. Hashmi, S. Balamuralitharan, and B. Bejena, "Optimal facial feature based emotional recognition using deep learning algorithm," *Computational Intelligence and Neuroscience*, 09 2022.

[50] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner,

W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, "Challenges in representation learning: A report on three machine learning contests," 2013.

[51] A. Lopes, E. Aguiar, A. De Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order," *Pattern Recognition*, vol. 61, 07 2016.

[52] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, 2010, pp. 94–101.

[53] N. D. Mehendale, "Facial emotion recognition using convolutional neural networks (ferc)," *SN Applied Sciences*, vol. 2, pp. 1–8, 2020.

[54] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-8, pp. 679 – 698, 12 1986.

[55] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1–10.

[56] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014.

[57] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Proceedings of the ... International Conference on Automatic Face and Gesture Recognition. IEEE International Conference on Automatic Face & Gesture Recognition*, vol. 28, no. 5, pp. 807–813, 2010.

[58] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *2005 IEEE International Conference on Multimedia and Expo*, vol. 2005, 08 2005, pp. 5 pp.–.

[59] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn, "Disfa: A spontaneous facial action intensity database," *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 151–160, 2013.

[60] T. Bänziger, M. Mortillaro, and K. Scherer, "Introducing the geneva multimodal expression corpus for experimental research on emotion perception," *Emotion (Washington, D.C.)*, vol. 12, pp. 1161–79, 11 2011.

[61] A. Dhall, R. Göcke, S. Lucey, and T. Gedeon, "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark," *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 2106–2112, 2011.

[62] I. Oztel, G. Yolcu, and C. Oz, "Performance comparison of transfer learning and training from scratch approaches for deep facial expression recognition," in *2019 4th International Conference on Computer Science and Engineering (UBMK)*, 2019, pp. 1–6.

[63] S. Palaniswamy and Suchitra, "A robust pose & illumination invariant emotion recognition from facial images using deep learning for human-machine interface," in *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, 2019, pp. 1–6.

[64] D. Lundqvist, A. Flykt, and A. Öhman, "The karolinska directed emotional faces - kdef, cd rom," CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998, iSBN 91-630-7164-9.

[65] M. Munsif, M. Ullah, B. Ahmad, M. Sajjad, and F. Alaya Cheikh, "Monitoring neurological disorder patients via deep learning based facial expressions analysis," in *Artificial Intelligence Applications and Innovations. AIAI 2022 IFIP WG 12.5 International Workshops*, 06 2022, pp. 412–423.

[66] S. Agrawal and P. Khatri, "Facial expression detection techniques: Based on viola and jones algorithm and principal component analysis," in *2015 Fifth International Conference on Advanced Computing & Communication Technologies*, 2015, pp. 108–112.

[67] A. Boughida, M. N. Kouahla, and Y. Lafifi, "A novel approach for facial expression recognition based on gabor filters and genetic algorithm," *Evolving Systems*, vol. 13, pp. 1–15, 07 2021.

[68] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 06 2014.

[69] Y. Tian, T. Kanade, and J. Cohn, "Recognizing action units for facial expression analysis. ieee transactions on pattern analysis and machine intelligence, 23(2): 97-115," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, pp. 97 – 115, 03 2001.

[70] N. Mehta and S. Jadhav, "Facial emotion recognition using log gabor filter and pca," in *2016 International Conference on Computing Communication Control and automation (ICCUBEA)*, 2016, pp. 1–5.

[71] F. Ahmed, E. Hossain, A. H. Bari, and A. Shihavuddin, "Compound local binary pattern (clbp) for robust facial expression recognition," in *2011 IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI)*, 2011, pp. 391–395.

[72] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0031320395000674

[73] D. Lakshmi and R. Ponnusamy, "Facial emotion recognition using modified hog and lbp features with deep stacked autoencoders," *Microprocessors and Microsystems*, vol. 82, p. 103834, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0141933121000144

[74] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, 2005, pp. 886–893 vol. 1.

[75] B. Deng, L.-W. Jin, L.-X. Zhen, J.-C. Huang, and H.-B. Deng, "A new facial expression recognition method based on local gabor filter bank and pca plus lda," *Information Technology - IT*, vol. 11, 01 2005.

[76] A. De, A. Saha, and M. Pal, "A human facial expression recognition model based on eigen face approach," *Procedia Computer Science*, vol. 45, pp. 282–289, 2015, international Conference on Advanced Computing Technologies and Applications (ICACTA). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050915003786

[77] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.

[78] S. M. Lajevardi and M. Lech, "Facial expression recognition from image sequences using optimized feature selection," in *2008 23rd International Conference Image and Vision Computing New Zealand*, 2008, pp. 1–6.

[79] A. Alreshidi and M. Ullah, "Facial emotion recognition using hybrid features," *Informatics*, vol. 7, p. 6, 02 2020.

[80] S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2584–2593, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:11413183

[81] Python Software Foundation, “Python,” https://www.python.org, accessed on 28-05-2023.

[82] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.

[83] Anaconda, Inc., “Anaconda,” https://www.anaconda.com, accessed on 28-05-2023.

[84] “Visual studio code,” https://code.visualstudio.com/, accessed on 28-05-2023.

[85] “Google colaboratory,” https://colab.research.google.com/, accessed on 28-05-2023.

[86] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, “Array programming with NumPy,” *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. [Online]. Available: https://doi.org/10.1038/s41586-020-2649-2

[87] T. pandas development team, “pandas-dev/pandas: Pandas,” Feb. 2020. [Online]. Available: https://doi.org/10.5281/zenodo.3509134

[88] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[89] F. Chollet *et al.*, “Keras,” https://keras.io, 2015.

[90] O. Contributors, “OpenCV-Python Tutorials,” *OpenCV documentation*, 2022. [Online]. Available: https://docs.opencv.org/4.7.0/

[91] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[92] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[93] T. Mei and Y. Rui, *Image Similarity*. Boston, MA: Springer US, 2009, pp. 1379–1384. [Online]. Available: https://doi.org/10.1007/978-0-387-39940-9_1014

[94] P. Y. Simard, Y. LeCun, and J. S. Denker, "Efficient pattern recognition using a new transformation distance," in *NIPS*, 1992. [Online]. Available: https://api.semanticscholar.org/CorpusID:11382731

[95] D. Huttenlocher, G. Klanderman, and W. Rucklidge, "Comparing images using the hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, 1993.

[96] P. Agrawal. (2022) A beginners' guide to image similarity using python. Accessed on 19 September 2023. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/03/a-beginners-guide-to-image-similarity-using-python/

[97] L. Wang, Y. Zhang, and J. Feng, "On the euclidean distance of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1334–1339, 2005.

[98] P. Gavrikov, "visualkeras," https://github.com/paulgavrikov/visualkeras, 2020.

[99] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 12 2014.

[100] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ser. ICML'13. JMLR.org, 2013, p. III–1139–III–1147.

[101] N. Keskar, J. Nocedal, P. Tang, D. Mudigere, and M. Smelyanskiy, "On large-batch training for deep learning: Generalization gap and sharp minima," 2017, 5th International Conference on Learning Representations, ICLR 2017 ; Conference date: 24-04-2017 Through 26-04-2017.

[102] I. Jammoussi, M. Ben Nasr, and M. Chtourou, "Facial expressions recognition through convolutional neural network and extreme learning machine," in *2020 17th International Multi-Conference on Systems, Signals and Devices (SSD)*, 2020, pp. 162–166.

[103] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine:

Theory and applications," *Neurocomputing*, vol. 70, no. 1, pp. 489–501, 2006, neural Networks. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231206000385

[104] Y. Zhou and B. E. Shi, "Action unit selective feature maps in deep networks for facial expression recognition," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 2031–2038.

[105] M. V. Zavarez, R. F. Berriel, and T. Oliveira-Santos, "Cross-database facial expression recognition based on fine-tuned deep convolutional network," in *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2017, pp. 405–412.

[106] M. Sari, A. Moussaoui, and A. Hadid, "A simple yet effective convolutional neural network model to classify facial expressions," in *Modelling and Implementation of Complex Systems*, S. Chikhi, A. Amine, A. Chaoui, D. E. Saidouni, and M. K. Kholladi, Eds. Springer International Publishing, 2021, pp. 188–202.

[107] K. Shan, J. Guo, W. You, D. Lu, and R. Bie, "Automatic facial expression recognition based on a deep convolutional-neural-network structure," in *2017 IEEE 15th International Conference on Software Engineering Research, Management and Applications (SERA)*, 2017, pp. 123–128.

[108] R. Melaugh, N. Siddique, S. Coleman, and P. Yogarajah, "Facial expression recognition on partial facial sections," in *2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2019, pp. 193–197.

[109] C. F. Liew and T. Yairi, "A comparison study of feature spaces and classification methods for facial expression recognition," in *2013 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2013, pp. 1294–1299.

[110] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua, "Fast keypoint recognition using random ferns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 448–461, 2010.

[111] S. K. Eng, H. Ali, A. Y. Cheah, and Y. F. Chong, "Facial expression recognition in jaffe and kdef datasets using histogram of oriented gradients and support vector machine," *IOP Conference Series: Materials Science and Engineering*, vol. 705, no. 1, p. 012031, nov 2019. [Online]. Available: https://dx.doi.org/10.1088/1757-899X/705/1/012031

[112] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, "Presentation and validation of the radboud faces database," *Cognition and Emotion*, vol. 24, no. 8, pp. 1377–1388, 2010. [Online]. Available: https://doi.org/10.1080/02699930903485076

[113] M. Wistuba, A. Rawat, and T. Pedapati, "A survey on neural architecture search," *arXiv preprint arXiv:1905.01392*, 2019.

[114] T. Elsken, J. H. Metzen, and F. Hutter, "Neural architecture search: A survey," *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 1997–2017, 2019.

[115] M. Chatzianastasis, L. Ilias, D. Askounis, and M. Vazirgiannis, "Neural architecture search with multimodal fusion methods for diagnosing dementia," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.